What contributes to the perception of musical phrases in western classical music?

Spiro, N.

Publication date
2007

Document Version
Final published version

Link to publication

Citation for published version (APA):
Spiro, N. (2007). *What contributes to the perception of musical phrases in western classical music?*. Institute for Logic, Language and Computation.

What contributes to the perception of musical phrases in western classical music?

Neta Spiro

What contributes to the perception of musical phrases in western classical music?

ILLC Dissertation Series DS-2007-02

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
phone: +31-20-525-6051
fax: +31-20-525-5206
email: illc@science.uva.nl
homepage: http://www.illc.uva.nl

What contributes to the perception of musical phrases in western classical music?

ACADEMISCH PROEFSCRHIFT

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof.mr. P.F. van der Heijden
ten overstaan van een door het college voor
promoties ingestelde commissie, in het openbaar
te verdedigen in de Aula der Universiteit
op woensdag 7 februari 2007, te 12.00 uur

door

Neta Spiro

geboren te Jeruzalem, Israël

Promotor:
Prof.dr. Rens Bod

Co-promotor:
Dr. Ian Cross

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

This research was supported by the Netherlands Organization for Scientific Research (NWO) in the context of the Innovation Impulse programme "Towards a Unifying Model of Linguistic, Musical and Visual Processing".

Printed and bound by PrintPartners Ipskamp.

ISBN-10: 90-5776-162-9
ISBN-13: 978-90-5776-162-1

Contents

Acknowledgements

The process of research and writing this thesis has involved enjoyable professional communication with numerous people of diverse fields and to all those mentioned below I would like to extend my warmest thanks.

To my promotor Rens Bod, thank you in particular for including me in the Innovation Impulse project "Towards a Unifying Model of Linguistic, Musical and Visual Processing". Thank you for inviting me to find my own way. This PhD research would not have been possible without you. Thank you to the Nederlandse Organisatie voor Wetenschappelijk Onderzoek for funding the research and the Institute for Logic, Language and Computation at the University of Amsterdam for hosting it.

To my co-promotor/supervisor Ian Cross, thank you for welcoming me so naturally into the Centre for Music and Science (CMS) community, for allowing me unlimited access to the new, state of the art and rare facilities (shown on page 40), and for giving me so much time and support over the years. Being one of the first to carry out research in this new and exciting laboratory has been a great source of excitement and honour.

Collaboration, discussion of statistical approaches, and technical support were the excuse for joyful communication with many. Beata Beigman Klebanov, working with you is always a pleasurable challenge. I hope we can continue to do so. Our ability to work together was much helped by Prof. Eli Shamir and Chana Slutzkin at the Selim and Rachel Benin School of Engineering and Computer Science at the Hebrew University, Jerusalem who welcomed me as a visiting scientist. Joel Swaine, many thanks for your technical assistance in preparing all the listeners' studies. Nick Collins and Dan Tidhar, thank you both for your help with programming, which has been instrumental to the project, and Elaine King, many thanks for the tempo tracking program all those years ago. Allan Pentecost and Vanessa Didelez, many thanks for your key statistical advice and Isabel Martinez, thank you for our conversations about click methods. Alan Blackwell, I am so glad we could continue our conversations throughout my time in Cambridge.

The CMS, Faculty of Music at the University of Cambridge has formed a focus. The old-timers Tommi, John, Jess, Matthew and Nikki, and Martin, Satinder, Letti and Ghofur thank you for the presentations, discussions, chats and Granta, forging our strong community. We have all struggled together and what a pleasure it has been. Thank you also to all of those who endured implicit and explicit phrase identification tasks - this work really could not have been carried out without you. Wagner will never forget you either.

Many thanks to Henkjan Honing, Colin Matheson, Rokus de Groot, Pieter Adriaans and Remko Scha for your interesting comments on the text from your individual perspectives.

Being attached to two institutions, there are so many individuals to thank. Frank Veltman, Ingrid van Loon, Marjan Veldhuisen, Anny Craje, and Jessica Pogorzelski many thanks for helping me with administrative questions at the ILLC, University of Amsterdam.

Many thanks to my officemates: Merlijn Sevenster for your practical enquiries and solutions, to Menno van Zaanen for our interesting chats and Yoav Seginer for your calming presence. Thanks in particular to Aline Honingh. We may have been distant for some time, but our experiences were close.

Beverly and John, thanks for the mulled wine. Matthijs Boekholdt and Suthesh Sivapalaratnam, we met in Cambridge, we meet in Amsterdam.

Over to Cambridge and thank you to Sue Round and Terry Wylie for welcoming me so warmly at the Faculty of Music, making me feel part of the crew from early on, and being so helpful at all times. Thanks to Trevor and David and I wish you both well in your retirement. Alasdair, Rebecca and Ag Parker, a home with an open fire and open arms.

At King's College, Cambridge, thank you so much to Nicholas Marston. Yet again, you played a pivotal role. Thank you to Stephen Alford, Bill Burgwinkle and Maria de la Riva for making me feel part of the college community so quickly. It has been such a pleasure. In particular, it has paved the way for meeting and re-meeting peers from distant and not so distant fields, including: Geert De Baere, thank you for translating the abstract and clicking the cover, Jonathan Mannering, Henry Mares, Delphine Mordey, François Soyer, and Owain Vaughan.

לאמא ולאבא תודה רבה וזה רק סימן

Neta, 2006

Cover
An allegory of musical phrases: their flow, journeys and boundaries,
the resulting arcs, expectations, memories.

View from Y11, King's College, Cambridge, UK

Abstract

The musical phrase

Is it? Who is it for? Where is it? What is it? When is it? Why is it?

This commonly used term, so intuitive to many musicians, has a variety of associations with the terminology of many disciplines including music, psychology, and linguistics. However, its nature remains obscure.

Is it?

One of the primary aims of this study was to establish to what extent there are common ideas about the nature of the phrase, its description, definition, identification and function.

Who is it for?

Another aim was to identify the types of population to whom this entity is relevant. This was done through investigating: 1) verbal and musical responses provided by listeners of different musical experience, 2) musicians annotating scores as if in preparation for performance, 3) performance characteristics of publicly available recordings, 4) discussions by music psychologists, music analysts and theorists, and 5) the musical analysis of pieces according to features. This comprehensive approach is referred to as 'the combined approach' below. These response groups have been studied before, though not with such direct and detailed methods. The results indicate that common aspects of the phrase are not learnt; listeners of different degrees of musical training or lacking it altogether responded similarly to phrasing tasks and questions.

Where is it?

This study discusses musical phrase examples of various musical genres and media. Some of these examples were previously investigated in the context of different disciplines. These range from folk to western classical music. However, the core of this study is the application of the combined approach to eight case-study pieces followed by analyses of seven test pieces all from the western classical repertoire.

One of the questions concerns the extent to which the phrases and their structures are clear 'from the score' (i.e. from the musical features that can be

identified in the score) and to what extent they are only clear in performance. By using responses to MIDI renditions (from listeners with different musical experience) and score annotations (by musicians), it is observed that common phrase structures are clear from their musical features. Musical features are musical elements that are combined and have particular characteristics in relation to their context. These include: cadences, relatively large pitch intervals, long notes or rests, repetitions, and changes in texture, motive, and harmonic rhythm.

Furthermore, the results show that there is also a rôle for performance features (changes in tempo and intensity). The identified tempo and intensity changes in recorded performances were also compared with: 1) phrase structure identified by analysts, score annotators and listeners to MIDI renditions, and 2) listeners' responses to the same performances. These, in turn, indicate that the same positions highlighted in performance are also identified in the other modes, and the listeners' responses to performances relate to the performance features. The main positions identified in response to performances and MIDI renditions are the same, but the proportions of responses differ. Furthermore, having heard one performance, its phrase structure seems to be remembered and affects the phrase identification of a subsequent performance (leaving "footprints"). Phrasing seems to be fundamentally 'in the music' and accentuated, clarified or obscured by performance.

Phrasing has mainly been discussed in the context of monophonic music. In this study, music of different textures is explored. The results of the combined approach indicate that in polyphonic music (e.g. melody and accompaniment) there may be differences between the phrase structures of different parts; phrases in different voices can lead to and complete each other, overlap or coincide, and these differences are often identified by participants. This indicates that we identify both the individual phrase parts and structures of the different musical voices, and accommodate these in a more general identification of conflicting, complementary or similar phrase structures.

What is it?

The term phrase has several related ones occurring in the literature and used by participants in the current study (including, segment, unit, chunk, sentence). Music psychological and computational approaches to grouping, segmentation and phrasing concentrate on the identification of boundaries, whilst other music theoretic approaches to phrasing discuss internal characteristics.

The results of the combined approach developed in this study (and described above), indicate that phrases include some of the following parts: beginning, beginning of the end (implication/expectation), end (initial arrival), prolongation (continuation of the end) and end of the end (end of the resolution), and that each one is indicated by specific musical features. Though all these different parts may be present, they do not have to be for the 'phrase' to be identified, recognised or implied. The above analyses also investigated the relative importance of different phrase parts, which determines the character of the phrase, such as front-heavy and end-heavy phrases, and possible relationships between phrases, such as, antecedent-consequent phrase pairs.

For all of these, the key seems to lie with the presence or absence of musical features. These belong to different categories, which have their different scopes of presence, impact and function. Some can be instantaneous (occurring, being identified and having their repercussion from one note to the next, such as a large pitch interval), some can be predictive (occurring over an area and creating expectations, including developing harmonic progressions such as cadences), and some can be retrospective (again occurring over an area but revealing their importance in retrospect, such as repetitions). Different features and feature combinations seem to systematically coincide with varying degrees of response, identified by the combined approach. Some features and their combinations are strongly indicative whilst others are less so. The former are rare whilst the latter are more common. Depending on the musical context (such as genre, instrumental combinations, or local context) common features acquire greater importance. Moreover, this combined approach highlights the interdependence of the musical features; different combinations of harmonic, metrical and pitch structure, for example, can form weaker or stronger phrase ends. The feature and phrase-part combinations can be such that more than one possibility can arise (sometimes resulting in 'ambiguity').

Though the relationships between the features, phrase parts and phrases are complex and depend on several parameters, they are formalised in a rule base. Unlike other rule bases the intention here is to reflect the process of phrase identification, including the 'weak' phrases, by participants, and providing alternative possibilities, using the concept of musical features developed on the basis of the combined approach. This rule base is formalised as an algorithm resulting in clear and consistent phrase structures, and may in future be implemented for the study of a larger corpus of music.

When is it?

These features and feature-combinations seem to result in candidate positions for phrase starts, ends and internal parts. Some positions are chosen by a majority of participants whilst others are less frequently identified. The latter coincide with weaker features and the respective weaker phrase parts. These would probably not be included in 'clean annotations' such as in the Essen Folk Song Collection, but seem to be an integral and important part of the processes of listening, performance, and analysis.

Moreover, through the combined approach discussed above and through the use of 'click' studies, unlike in previous psychological studies, it is here revealed that phrase parts are often identified over a period rather than on specific notes.

Why is it?

This study indicates that the phrase is both an organising and organised unit (in this way similar to a linguistic sentence) that is related to memory, breath, and physical motion. It gives structure, framework, order and reference, and interacts with other structures of different types (such as, metrical structure). Its length is often described as constant. However, the results of the combined approach indicate that there is great variability in phrase length. The identification of these units may contribute both to recollection and comparison between similar phrases and to the more general structuring and memory of the music. The phrase helps in following motion or progressions from a beginning to arriving at a destination or returning. Musical implications, and therefore expectations, seem to play an important rôle in this progression. Moreover, from the way in which it, its musical features and characteristics are used, and their frequent occurrence in discussions of music analysis, performance and perception, the phrase seems to be essential to our capacity to follow the kind of music studied here.

Chapter 1

Introduction and Background:
An Elongated Up-Beat[1]

1.1 Introduction

General questions

How do we follow music? How is this related to following other strings of information? Is the general need to detect order in our surroundings an attempt to help us understand them? Can the idea of the phrase help explain how we do so? Perhaps the musical phrase is one manifestation of our seemingly unending need to find structure in incoming information. It may reflect our tendency to guess (generate expectations of) what will be next, to be surprised by deviation from these expectations, and to keep guessing the next development – an exciting, insatiable quest.

The term 'musical phrase' is often used and relied on in discussions in various musical contexts, including: music theory, analysis, performance, psychology and computational approaches to music. However, explicit definitions of the term are rare, indicating that there is a common understanding and acceptance of its meaning. In general, the term seems to imply a unit of music that has an identifiable beginning and end, one that is self-contained but has within- and between-phrase structural characteristics. It is almost always connected with its 'linguistic roots', bringing to mind both 'grammatical' characteristics associated with construction and 'practical' concerns associated with breath and expression.

The musical phrase seems to rely on subdivision according to the whole musical "sound" rather than being limited solely to one element of the musical surface. It therefore refers to musical entities that can have a range of musical characteristics. This range is so broad and the variety of emphasis placed on different musical parameters by different writers having their distinct standpoints is so great, that there does not seem to be a consistent definition of the term. However, it seems

[1] Rothstein (1989)

that the common characteristics and the wide-ranging implications of the different definitions require a comprehensive exploration of the different musical features, their interrelation and the ways in which they contribute to the perception of music. Musical features are musical elements that individually or in combination have particular characteristics in relation to their context. These include: cadential and voice-leading progressions, relatively large pitch intervals, long notes or rests, repetitions, and changes in texture, motive, and harmonic rhythm.

It seems paradoxical that on the one hand musical phrasing is regarded as essential for composition, analysis, performance, listening and perception, whilst on the other hand, the discussion of theory and practice of the definition, meaning and use of musical phrases is surprisingly small.

Definition of questions

This paradox has many facets. Phrasing seems to have different definitions, terminologies and meanings for different components of music composition, analysis, performance, listening and perception. To some extent, the differences are those of emphasis. Some music analysts base their theories on those put forward in composition manuals, but then develop additional aspects, primarily longer-term harmonic concerns. Music psychological theories, on which many of the computational studies are based, have a different starting point, that of general psychological characteristics. These psychological characteristics are interpreted in terms of musical elements, primarily those at phrase boundaries, which are not emphasised to such an extent by music theorists. The few examples of performers' writings indicate that they have different concerns, ranging from the broad analogies with breath to the specifics of articulation related to phrasing. Analyses of performers' recordings however, indicate that elements discussed by music theorists may also be important for their phrasing, as performance characteristics coincide with locations identified by theorists. There have been few studies of listeners' responses to phrasing. The key studies either modify examples to test specific musical elements, or assess the effects of psychologically related musical features.

Another facet of the paradox is the scope of the applicability of the musical phrase: whether it is limited to a theoretical description of musical structure, or whether it is also used when preparing for performance and listening to music. Yet another facet is related to musical experience; if phrase structure is indeed operational in perception, does it have different definitions, terminologies and meanings for those with different musical experience, is it limited only to the musically very experienced, or does it apply commonly to all?

A potential paradox may arise from a duality of organisation and its disruption. On the one hand, phrases help to organise the music by contributing to the understanding of its structure. Not only is there a retrospective understanding of the structure, but also there are expectations of what will happen next. On the other hand, these expectations are not always fulfilled, keeping us interested, and there may be several different structural interpretations of the same music. The musical phrase provides a framework with a range of degrees of organisation from strict to loose.

The wider picture

Although here phrasing is treated primarily in its structural dimension, it may also contribute to other aspects of the musical experience. For example, emotional responses to music are attributed to extra-musical sources but also to musical ones. One of the main musical sources is thought to be the expectations generated, their disruption or their fulfilment. In this way, the investigation of phrase structure and its understanding may contribute to the understanding of some emotional (or motional) responses to music.

Many attempts to describe and understand this primarily non-verbal art form have been to do so by analogy: physical and verbal. For example, the trajectories of pieces of music have been compared to the progression of physical motion, including ideas of movement from one place to another (from a start to a goal). Another one is the change in pace of that movement, acceleration and deceleration, even to the extent of comparisons with the exact rate at which objects and people decelerate. Both of these motion analogies also arise in discussions of phrases – the movement towards a goal, and the rate of deceleration at the end of the phrase. The structure and, more broadly, the function of music have been compared to language, as a form of communication, as a 'rule based' structure, and as a physical experience involving, for example, breath. The details of analogies and comparisons that arise specifically in discussions of the musical phrase are quite extensive: the comparison of phrases to linguistic sentences, the grammatical structure dictating phrase structure, the need of breath defining the length of the phrase and more broadly, phrasing helping to reveal and clarify the structure and therefore its communication to the listener.

The ideas of organisation into units that help us both to understand the incoming information and to remain attentive, by introducing the interest generating deviations from our expectations, are often discussed in other areas that involve information processing. As in language, literature, art and architecture, these follow basic psychological principles. Within these fields, especially psychology, a number of terms have been used to describe units of information including segmentation, chunking and grouping. Each one has its own specific connotations while having certain commonalities, especially the ideas of units and subdivision of a larger whole.

The term musical phrase is investigated in this study for several reasons. 1. It is important to determine the extent to which the musical phrase is indeed similar to these other units of information before any of those terms are used. 2. This study, though using psychological approaches, is strongly focussed on the musical elements, and their rôles. 3. The musical phrase despite, or because of, the paradoxes described above may already have a sound basis for a definition and meaning. 4. Other terms have their own associations, which do not seem suitable for music. It may be that the commonalities between the units in different domains are not those that are basic to these terms: for example, a segment is usually considered to be a part or a subdivision of a whole, while a phrase is a constituent that builds that whole by relating closely to the other phrases. A chunk has associations with a homogeneous unit, while phrases are not homogenous; they have internal structures made up of different components. Groups, like the other two terms, can refer to units of any size, from the very small to the very large, and it seems that many different types of units can fall under the term group. In general, it seems that the parts of the units that are concentrated on in discussions of these three terms are the boundaries and the hierarchical relationships between them. The musical phrase, however, may include an internal structure and a dynamic aspect, musical elements leading to a goal within the phrase while a sequence of phrases may lead on.

Approaching the questions

This study approaches the above questions from two broad perspectives: 1) a study of the literature on the subject to investigate what the phrase means to practicing musicians and theoreticians from the different musical domains, and 2) empirical investigations of people's responses to questions and tasks of phrase identification and definition, parts of which are related to studies explored in (1) and parts of which are new approaches.

Previously the phrase has been investigated or described in broadly four domains: music theory and analysis, music performance, music psychology and computational approaches to music.

Music theory and analysis

In the music theoretical and analytical domains there are broadly two types of sources of information about the phrase. One that takes the phrase as the main topic of investigation and outlines its characteristics in relative detail, and the other that takes the phrase more or less as a given, and uses the term in the description or analysis of pieces of music to different degrees of specificity.

There are only a few examples of writings from the 18th and 19th centuries that concentrate on musical phrases. For example, Reimann (1884) advanced a model of phrase structure that he believed to be a constituent of all classical music. The basic building blocks are: 1) Taktmotive which are musical segments that contain only one strong beat (e.g. 1 bar) that may be preceded or followed by weaker beats. 2) Taktgruppe which are segments consisting of two Taktmotiven combined into a unit in which the centre of gravity is the second bar. 3)

Halbsätze, half phrases which are segments consisting of four bars, with the centre of gravity in the fourth bar. 4) Perioden, which are segments combining a Vordersatz (antecedent phrase) and Nachsatz (consequent phrase) and forming a unit of 8 strong beats, resulting in eight-bar (or rather twice four-bar) unit that was supposed to be the universal model of musical organisation. However, the inflexibility of this square design prevented the accommodation of the many phrases containing an odd number of bars, frequent even in the Classical repertoire, and the pervasive irregularity of Baroque melodic designs (Neumann, 1993). The ideas of phrases having an even number of bars and a range of phrase lengths, and the relationship between antecedent and consequent phrases have, nevertheless, remained common in phrase descriptions and definitions.

Heinrich Christoph Koch (1787, 1983) dedicated a section of his composition treatise to phrases. He describes a number of different types of musical units, primarily defined by the degree of (mainly harmonic) close at their end. Like Reimann (1884), he also discusses phrase lengths, emphasising the four bar phrase but also suggesting up to seven bars for a 'basic' phrase. He suggests that equal length phrases are preferable. However, he then goes on to discuss how these basic phrases can be extended and combined. Like many theorists in all the different domains, he also makes general and specific comparisons with language. Koch attempts to identify subject – predicate units in music but finds that this is not possible. Despite the in-depth description, Koch points out that, in the end, 'feelings' are needed for the identification of phrases.

William Rothstein (1989) based his theory of phrase rhythm in tonal music partly on Koch's work. For him, the most important aspect of phrasing is complete tonal motion. This is not only reflected in the 'vertical' harmonies, but also in the voice-leading (inspired by, for example, Schenker, 1979). Rothstein discusses the different types of phrase endings (primarily harmonic, but sometimes more long-term than Koch), the hierarchical relationships and the relation between phrases (such as antecedent and consequent), the lengths of phrases (the relative length of different phrasing causing 'phrase rhythm') and the preference for phrases of equal length. He describes how the basic structure and length is modified by elision or expansion of different types. This seems to provide the possibility of identifying a basic phrase structure and its modifications. Rothstein also distinguishes between metre (and hypermetre) and phrase structure. He discusses the tension between the two and how they can sometimes strengthen each other.

These music theorists begin with the description of a basic phrase. In this context, they describe length and harmonic structure (both local and more long term) and mention voice-leading principles. Some theorists, when describing phrases in specific pieces, take up Rothstein's criterion for complete tonal motion, whilst others, identify phrases that do not necessarily end with complete tonal motion (such as Temperley, 2001). In this study, the extents to which the ideas of complete tonal motion (mainly identified through cadences) and phrase length are investigated.

Music performance

In their discussion of phrasing performers tend to emphasise the connection between language and music, concentrating on the similarities with speech and clarification of the musical structure. This approach differs from that of music analysts and music psychologists in that the latter concentrate on the rule base aspects of musical structure. In their discussions, performers rarely point out specific musical elements that contribute to phrasing decisions but instead focus on how to communicate the phrase and why this is necessary. They often also focus on the difference between articulation (the very local details of performance) and phrasing. The relative lack of material written by performers (and composers) on the subject of phrasing and in particular specific musical elements that influence their phrasing decisions can be overcome, to some extent, by the analysis of performances.

Analysis of music performance often concentrates on tempo and dynamics and their changes. These studies often describe these changes with respect to phrase structure with *ritardandi* ('phrase-final lengthening'), and *diminuendi* often being associated with the ends of phrases (Todd 1985, Shaffer and Todd 1987, Clarke 1988, Repp 1990). The greater the changes in these two characteristics, the 'more important' the phrase boundaries are. Some also describe an *accelerando* and *crescendo* at the start of phrases, creating together an 'arc' within the phrase. This study investigates the extent to which it is possible to use these performance characteristics to identify phrases, and the elements that contribute to their identification and perception.

Music psychology and computational approaches to music

In music psychology, as in music theory and analysis, the term phrase appears in two types of sources: studies that investigate the phrase and those that use it as part of other investigations. There are also two types of approaches: empirical (Deliège, 1987; 1998; Palmer and Krumhansl, 1987a; 1987b) and theoretical (such as, Lerdahl and Jackendoff, 1987). The empirical studies of phrasing have been of three types: 1) The investigation of the contribution of a small number of musical elements to phrase perception. This is achieved by constructing musical examples in which the same basic musical material is presented to listeners in the form of several different variations (Palmer and Krumhansl, 1987b). 2) The investigation of responses to a recorded performance of a piece from the published repertoire by asking listeners to identify phrases. This is done in order to compare the responses of groups with different levels of musical experience, and of those with different levels of familiarity with the piece, and to explain some of these responses in relation to the presence of Gestalt based musical elements (Deliège, 1998). 3) A 'click detection' method has been used to investigate the exact location of 'phrase boundaries' (such as Stoffer, 1985). This method is based on the idea that when we are processing information within a unit, the cognitive processing load is high and, conversely, when we are between units the cognitive load is low. When the cognitive processing load is high, i.e. when we are processing something that is the middle of a unit, then two behavioural consequences are expected: we react relatively slowly to an external stimulus that is not related to the unit we are processing, and in our memory that external stimulus is remembered as having occurred after the end of the unit, and not during it. Conversely, when the cognitive processing load is low, i.e. between the units, we react more quickly to the external stimulus and we remember its position correctly. Although some studies used this method for the exploration of the location of phrase boundaries, there is some debate about whether the examples used really reflect 'phrases' as identified in many theoretical and empirical works. In this study, aspects and ideas of each of these approaches are explored further and combined with other approaches.

The theoretical studies are often based on psychological, Gestalt principles. They often put forward rule bases, sometimes developed as far as 'models' that are ready for a computational implementation to a greater (Temperley, 2001) or lesser (Lerdahl and Jackendoff, 1987) degree. Usually, the more computational the study, the smaller subset of musical elements is. In these approaches the emphasis is on finding systematic rules that can lead to the identification of groups, or rather group boundaries in (until now usually) monophonic music.

Another computational approach consists of the collection of 'rules' directly from a large corpus which has been manually annotated with musical phrases, thus representing the exact memory of previously heard phrases. These rules are then used in interpretation of new music (Bod, 2002).

Some studies investigated phrasing directly (such as, Palmer and Krumhansl, 1987a; 1987b; Temperley, 2001) while others included phrasing as part of a broader category, for example segmentation (Ferrand et al., 2003) or grouping (Deliège, 1987; 1998). Grouping approaches usually consider each note as part of a collection of notes around it. Each collection of notes is related to the others around it in a hierarchical manner, and the group size grows depending on the level within the hierarchy. The rules governing these groupings have been inspired by either gestalt-based psychological principles expressed as a rule base, or by memory-based approaches represented as databases treated statistically. Both of these keep a close relationship to language perception studies in, for example, constructing tree-structures to represent the grouping structure. It seems that the definitions of a 'group' are also based on the rules that are used to identify them, mainly focussing on elements at the group boundaries such as temporal gaps (rests and relatively long notes) and pitch gaps (large intervals), or repetition of previous material. As in music theory and analysis, psychologically based studies of phrasing often relate musical phrasing to the grammatical structure of language, both in the reasons for it and in the idea of the rule base governing it.

The musical elements concentrated on in psychological approaches include temporal gaps, both rests and relatively long notes (Lerdahl and Jackendoff, 1987; Temperley, 2001), relatively large pitch intervals (Lerdahl and Jackendoff, 1987), repetition (Cambouropoulos, 2001) and phrase length (Temperley, 2001).

The approach used in this study incorporates elements from each of the above approaches, explores other empirical and analytical methods, and investigates a broader range of musical and psychologically-based elements than in previous individual studies. Furthermore, this study investigates a wide range of musical examples from various eras and genres from the western classical repertoire using a selection of case-study pieces. This is referred to as the "combined approach".

Perception is a very comprehensive term and a definition relevant to the present study is that of Matlin; the use 'of previous knowledge to gather and interpret the stimuli registered by the senses' (Matlin, 2003, pp. 32 and 500). This study investigates the markers that may provide order throughout the many types of our musical experience and, more specifically, the musical features that we gather through this experience, and the results of using this information. This study therefore investigates the 'perception' of phrasing.

Signposts and junctions

The overall aim of this study is to establish the relationship between musical elements and phrasing, and uses a number of techniques to view and analyse this relationship from a wide range of perspectives using the combined approach.

More specifically, the study aims to investigate the extent to which identification of different phrase structures relates to various musical elements and their combinations, and whether these are affected by the form of the musical renditions and the experience of the listener, performer or theoretician. From these general musical elements, musical features are initially identified from general principles of music theory and the literature of music theory, analysis, psychology and computation. The main emphasis of the empirical part of the work is to explore the use and effects of these features on the identification of phrases in pieces from different genres in western classical music taken as case-studies. This kind of approach allows the exploration of the subject matter in a systematic way, whilst allowing for further investigation on the basis of the acquired results and the evolving hypotheses.

In so doing, the study also arrives at the identification, description and analysis of phrase–parts and investigates the extent to which these are useful in characterising internal phrase structures. The investigation of the internal structure of the phrase and phrase parts leads to the suggestion that expectation**,** especially of the phrase end, is an integral part of phrase identification (an aspect that to my knowledge has not been considered in depth in psychological and computational approaches to phrase perception so far). This approach to the musical features and phrase structure helps explain the phenomena of continuous "interest" during listening to pieces of music.

This description of phrases, phrase-parts and the expectations generated, also allows the exploration of areas that could be considered as ambiguous. The particular musical features and their combinations that render parts of phrases to be considered ambiguous contribute to an understanding of the way in which listeners' interests are generated and maintained.

The question of the universality of phrasing, whether or not different listeners identify the same phrases and to what extent they have the same perception, is also investigated. It is currently assumed in most music-psychological and computational studies that different listeners have similar phrase interpretations. If multiple phrase identifications are found to occur, a second question that follows is; what are the reasons for and the characteristics of multiple phrase identifications? This study investigates these questions from a number of different perspectives. Firstly, the study tests whether phrasing is considered differently when it is decided 'online' during listening, when it is decided in retrospect, or when the music is analysed through playing and/or studying the score. It then investigates whether there are basic commonalities among all three approaches. Furthermore, the study assesses whether or not there is greater variety of phrase identification during the 'online' listening than in any other approach. It further investigates the features contributing to ambiguity and whether a more detailed study of ambiguous areas contributes to the clarification of the phrase structure. This leads to the study of the effect of different degrees of emphasis of musical features contributing to strength of ambiguity in performance.

The study also investigates whether retrospective interpretation (and longer-term knowledge of the music, even by ear) may result in longer phrases and in more consistent identification of the phrase structure among listeners. Moreover, it investigates whether these possibilities arise because the listeners respond to different musical features, some stronger, encouraging more agreement among listeners, and some weaker, where differences between interpretations are more common. In addition, this study investigates to what extent phrase perception is affected by the listener's general musical experience and familiarity with the specific piece.

Analytical approaches

1. The empirical part of the investigation consists of a number of studies exploring the above questions from a number of different perspectives. Each study begins with the analysis of musical scores. Scores are the primary source of the western classical music. They provide the possibility of repeated analysis, the opportunity to compare parts of the piece and "travel back and forth through the piece".

There are two approaches to the analysis of the musical scores: 1) using and comparing existing analyses from the literature, and 2) carrying out musical analyses based on general music analytical principles (including motivic such as Rétian analysis and structural such as Schenkerian analysis). The analyses of the case-study pieces in the literature sometimes specifically discuss phrases but more often use the term within their analysis without defining it. Some of the studies mention or discuss specific musical features that lead to phrase identifications at various levels, such as only the location of phrase starts or ends, or areas that are problematic. All these types are used in this approach. For some case-study pieces phrase aspects are not discussed in the literature and so only the general musical analysis can be used. A set of features, phrases and phrase-parts for each piece are identified in these analyses. This approach is used partly in order to avoid the need of constructing simplified aural examples that do not contain the usual relationships encountered in music.

2. The next step is to compare the features and feature combinations to a number of different kinds of interpretations gathered through different types of empirical studies, including studies of performance characteristics, listeners' and performers' responses.

2.1. Initial exploration: Phrasing in songs

The introductory study is based on songs from the 19th century Lieder tradition and opens up several different topics of discussion: 1. The identification of phrases by performers. 2. The relationship between words and music 3. The agreement among performers and music analysts. 4. An initial exploration of the relationship between musical elements and the phrases identified (chapter 2).

2.2. Listeners' responses to MIDI renditions of range of case-study pieces

There are several specific aims for each study in this work but one of the aims common to all of them is to investigate the extent and the nature of the relationship between phrase identification and the musical features identified in the musical analyses. To this end, in this study, listeners were asked to identify phrases and positions at which they began to expect the end of the phrase while listening to the MIDI (Musical Instrument Digital Interface) renditions. The MIDI renditions provide 'dead-pan' performances, presenting only note-length and note-pitch directly from the score and without 'performance features' such as changes in dynamics and in tempo. This allows the 'musical features' to be studied more directly than from responses to recorded performances which include performance features (chapter 3).

2.3. Listeners' responses to performances of the same case-study pieces

A further experiment investigated phrase identification in different performances of the case-study pieces. The aims included the assessment of: 1) The extent to which the phrase identification is different from those obtained for the MIDI renditions. As part of this experiment therefore, in addition to the new listeners, some listeners from the MIDI experiments were asked to return for further sessions of listening to performances. 2) The extent to which different performances (or performance characteristics) have an effect on the phrase identification. As in the MIDI experiments, listeners were also asked to identify the positions at which they began to expect the end of the phrase. Again, the results for both tasks were compared with the features identified in the music analyses.

There may be the impression that musical phrasing is decided only on the basis of performance features (such as breath, dynamics and tempo). In this case, the investigation of the phrase should be based purely on performance features. This study explores to what extent phrases can be identified without performance features and then what the additional effects of performance features are. This is followed by a comparison of listeners' responses to recordings by several performers having different performance characteristics, emphasising different musical features in the same case-study pieces.

For both listening experiments a number of factors that are not directly related to the musical features of individual pieces are investigated, such as: the effect of musical experience of the listener, previous familiarity with the piece, 'learning' of phrase structure within the listening sessions, and the effect of hearing different performances of the same piece.

2.4. Listeners' phrasing study – A statistical method

For both listening experiments, an alternative method of analysis is explored which aims to asses statistically both within-person and between person consistency. During the application of this method several questions that arise in any analysis of this kind of data are discussed. These include the most useful or meaningful temporal unit for presentation of the data and the categorisation of responses as similar or different. Proposed solutions, some provided by this method, are then explored (chapter 4).

2.5. Location of the 'boundary' - Click detection study

Having explored various types of studies of phrase identification, this one is carried out in order to explore the more specific location of phrase ends and starts. This method has been applied in both language and music perception studies before, and relies on theories of cognitive load (chapter 5).

2.6. Musicians' phrase notation through playing from the score

The next study returns to the method used in the introductory study (see section 2.1), this time with some of the pieces used in the listening studies. This is primarily to compare the listeners' and musicians' phrase annotations (chapter 6).

2.7. Analysis of performance characteristics in recorded performances

A different approach to the investigation of phrase identification and its comparison with the musical features identified in the musical analyses is the study of recordings of different 'master' performances of the same piece. This study tests whether the musical elements are reflected in the performance features. The study also investigates the similarities and differences between performance characteristics of different performances (chapter 7).

2.8. Music analysis of case-study pieces

Following the presentation of the above studies, the different case-study pieces are analysed in a number of different ways. Firstly, different approaches discussing phrasing or related structures such as groups and segments are discussed and, where possible, applied to the current case-study pieces in order to: 1) review in detail the current definitions of, and assumptions about, the term 'phrase', 2) identify the cues and explain decisions of phrase identification and definition, 3) investigate the results of these studies in light of their underlying theories and evaluate their general applicability, and 4) interpret results of the current study using the various theories and rules given in these studies (chapter 8). This is followed by a discussion of music analysis and a detailed presentation of two of the most developed music analytic theories of phrasing (by Koch and Rothstein, chapter 9). These two chapters and the previous studies prepare for the analysis of the case-study pieces and the comparison of these analyses with the results of the current phrasing studies (chapter 10).

2.9. Musical features and phrases

The analysis of the case-study pieces from all the above perspectives allows a quantitative assessment of the relative importance of features and their combinations as phrase-part indicators, and the relationship between types of features and the phrase-parts and phrase-types with which they occur (chapter 11). This leads to a comprehensive discussion of musical features and their rôle in phrasing (chapter 12). This is followed by a comprehensive discussion of phrasing in analysis, performance and perception, including a discussion of phrase-type categories and internal structures of phrases (chapter 13).

2.10. Rule base

On the basis of the combined approach, though the relationships between the features, phrase parts and phrases are complex and depend on several parameters, a rule base is formalised. This rule base is designed to reflect the process of phrase identification by listeners, musicians, performers and analysts. It includes the "strong" and "weak" phrases and provides alternative possibilities. This rule base is also presented as an algorithm, which may in future be implemented for the study of a larger corpus of music (chapter 14).

2.11. Test pieces

The rule base in particular, and the ideas developed through this combined approach in general, are re-examined using a group of seven test-pieces. Tempo and intensity contours of performances of these pieces are analysed and an expert analyst provided his score-based interpretation of the phrasing. The results are compared with the analysis according to the musical features and the implied phrase-parts as formalised in the rule base (chapter 15).

Apparent hurdles

1. Many of the previous studies, have drawn wide conclusions on the basis of a very small number of musical examples, sometimes only one piece, using one technique, and sometimes one population of listeners. The empirical parts of this study, though based on a small number of musical examples, study eight pieces in great detail. This might still be seen as potentially leading to over-generalisation on the basis of limited sources, but it is a broader set than in any of the previous empirical studies of western classical music. Furthermore, this potential limitation is countered by: 1) reference to and analysis of other examples, 2) reference to published material on the subject which uses both the same and other pieces as examples, and 3) the employment of the findings of the current study in the analysis of the seven additional test-pieces, in comparison with the responses of the expert analyst and performance contours.

2. Some previous studies use musical examples that are either newly constructed for the specific task, substantially modified versions of a great classic piece (Palmer and Krumhansl, 1987a; 1987b), or pieces from the repertoire (Cambouropoulos, 2001; 2003; Deliège, 1998; Ferrand et al., 2002; 2003). As there is such a lack of clarity as to whether there is agreement about phrase identification in pieces and what the reasons are for the identification of one phrase or another, it seems to be premature to begin with the construction of new musical examples. Instead, the combination of musical analysis of existing works from the repertoire with the empirical results seems to be a genuine representation of the way phrases are perceived and allows systematic way of investigating the musical phrase.

3. As in most psychological studies, the number of participants is a limiting factor but the population size here is not smaller than in most previous phrasing studies and moreover, here this is partly countered by the number of variety of tests applied to the same question and population.

The application of the combined approach to the study of phrasing should yield a comprehensive insight into the nature of the musical phrase and its rôle in our perception of western classical music. This, in turn, should reveal aspects of the way we treat information, and maintain interest and enjoyment.

1.2 Background

1.2.1 What is a musical phrase?

This study begins the exploration of the notion of a musical phrase with a review of: its meanings; in which disciplines it has been explored and what theoretical basis these can provide; how phrases have been defined and identified; the rôle of phrases in composition, performance and listening; and which musical characteristics contribute to their structure and perception.

Many aspects of the musical phrase have been investigated in the purely musicological context whilst others, related to the listeners' reactions, communication, perception and emotion, have been included in the psychological context of music cognition.

Several ideas derived from the different disciplines are included in descriptions and definitions of the phrase. One of the prominent ones is that of 'function' by structure giving and clarification. Another is related to linguistics: ideas encapsulated in rule bases are important in determining the phrase structure. Definitions and descriptions also often include ideas expressed as metaphors: the phrase is compared to breath or is described as having a directed motion towards a goal, controlled by harmonic motion and usually a cadence. A phrase is sometimes said to contain an element of expectation, which may or may not be resolved. Its structural characteristics may include an 'ideal' length, a range of lengths, or length relationships between the phrases. Rhythm has been related to these length relationships (phrase rhythm) and has been described as a within phrase characteristic. Phrases are not isolated units and the relationships between phrases, both adjacent (such as antecedent-consequent) and hierarchical, are often described. Phrase descriptions often highlight musical elements that may be involved in the construction or identification of musical phrases. These ideas varied over time, being affected by the contemporary views in philosophy and the other arts. This background chapter presents and discusses a number of these aspects.

Early theorists

Theories of musical phrasing have been developed over the centuries and influenced by several disciplines. In the 17th century, they were developed out of rhythmic theory and conceived in terms of poetic metrical theories. In the 18th century, the rhetorical analogy of punctuation and parts of the human body was introduced; Couperin (1772) used it as justification for the comma. Mattheson (1737) compared phrasing with parts of the human body. Concentration was on the 'anatomy' of phrasing: the identification of phrase starts and ends (Neumann, 1993, p. 272).

Riemann (1903) suggested that phrase structure is generated by processes of linear growth rather than by abstract patterns of stressed and unstressed units. He developed a precise notation for phrasing in which the piece is related to a theoretical eight-bar structure with a system showing modifications.

Though Riemann's influence was strong, his views were challenged by, for example, the Urtext movement, the virtues of which were explained by Schenker (1925, 1994) who saw no difference between legato (articulation) slurs in conventional notation and the slurs he used in his analytical graphs. These were conceived in terms of performance, and were also intended for study by performers. 'Despite Schenker's clear interest in performance and, in particular, in articulation and phrasing, the subject remains undeveloped in his theory, and has not even yet been fully integrated into theory' (Chew).

Early scores

More generally, the phrase arc, as a marking in the score, is a common idea. Composers, including Mozart (1756–1791), continued to show interest in the precise notation of articulation and attempted to refine it. Theorists, since the 18th century at least, have proposed ways of systematically marking phrases on the score (crosses and circles, such as Schulz, 1771, different types of strokes or beam connections such as Bach, Türk, 1789, Kirkpatrick, 1984, Reimann 1903 for vocal and instrumental music) but their ideas were not realised in a systematic, long-lasting manner. Well into the 19th century both theory and notation remained far from rigorous because of the common norms of performance and ways of communicating phrasing and other conventions (Keller, 1965). Moreover, though there are often arcs in scores, they are not a systematic marking specific to 'phrases' as they also indicate, for example, bowing or local articulation.

Historical development: by era or by composer

Phrasing is often described using general statements such as: music from the Medieval and Renaissance eras is pre-phrasing; Baroque music does not have phrases but rather either shorter units (motives) or larger sections inspired by rhetorical structure or dance, which have symmetrical designs, often eight-bar groups (Neumann, 1993); phrasing of Classical music is four-square, symmetrically and hierarchically clearly organised; Romantic music began by pulling at the extremes of phrase structure and, by the end of the era, broke down completely with Wagner (see also, Salzer, 1987, p. 8). Another view is that individual composers have their own characteristic phrase structures (such as described by, for example, Keller, 1965; Neumann, 1993).

However, there seems to be great variety both within eras and composer's oeuvres, while at the same time other aspects seem to be common across both eras and composers. Musical features and phrase types seem to be used in different combinations across the eras and with great variety with each era, composer, genre and even piece. Although individual characteristics of composers, eras, or genres should not be ignored altogether, such broad-brush statements therefore seem to obscure both the commonalities and differences.

This short historical overview indicates the changing concern of composers and theorists about phrasing and the importance they attached to it. It shows the origins of the surviving remnants of these attempts and the difficulties in establishing a systematic code for vocal and instrumental music; though the phrase seems to have been a preoccupation, it has been difficult to transform it into systematic markings in scores. It shows that:

- During much of the history of western classical music, 'the phrase' has been important.
- There were attempts to codify phrase notation, some of which are still used.
- Phrasing ideas varied, affected by the nature of the music, performance needs and current theories in other disciplines.
- There are a number of common recurring ideas such as; music consists of sections of different size and importance and a phrase is one type of section.
- Musical phrasing has been regarded as analogous to sections of language such as couplet or sentence controlled by several factors including breath.

This background section prepares for subsequent chapters by presenting ideas, approaches and methods of previous studies, broadly in their order in the following chapters and aspects summarised above are included in the individual areas of this study.

1.2.2 Phrasing in songs

The term phrase seems to have been first associated with vocal music and often even the shortest descriptions of the musical phrase include the importance of breath (both physical and metaphorical, section 1.2.10.5.1) and the comparison with linguistic structures. In this section, general comparisons between music and language, and text setting are discussed.

Linguistic connections

Authors make numerous and diverse types of comparison and analogy between music and language. For example, Chopin's comparison of music and language is among the more general; "'He who phrases incorrectly is like a man who does not understand the language he speaks'" (Keller, 1965, p. 4).[2] Others mention punctuation marks (Keller, 1965; Riemann, 1884) or the way a piece of prose,

---

[2] 'A term adopted from linguistic syntax…The term 'phrasing' implies a linguistic or syntactic analogy, and since the 18th century this analogy has constantly been invoked in discussing the grouping of successive notes, especially in melodies' (Grove: Phrase definition).

poem or speech, is organised in a hierarchy of units of different structural levels (Neumann, 1993, p. 260). Inspired by theories of natural language processing, some, including Lerdahl and Jackendoff (1987), compare the perception of musical structures to that of linguistic grammars.

Text-setting in songs

On the one hand, the text of a song is an 'external', non-musical cue that, at least theoretically, forms the basis for the structure of pieces. Therefore, in the context of western classical music, songs may be among the most constrained in their phrasing. On the other hand, there can be conflict between the structure of the music and text and the 'solution' depends on the context (Barra, 1983, p. 35).

In preparing a performance, according to Stein and Spillman, performers must first study the poetry, then the performance problems, and then each aspect of the musical structure in turn, recombining these steps in performance (1996, p. xiii). They give a detailed account of the different poetic structures and their relation with musical ones (1996, p. 334). They base most of their discussion of phrasing on that of Rothstein (see chapter 9). For them, phrases are composed of small rhythmic motives that combine to create larger musical phrases, which, in turn, combine to create entire musical sections (1996, p. 167). Discussing the phrase norms of 19th century Lieder, they explain that, although many theorists consider the eight-bar phrase to be a norm, they consider it to be four or, in slow works, two bars. These can usually be sung in one breath, and the even number creates a sense of symmetry and balance. Other norms include the antecedent-consequent structure, reinforced by norms of poetic texts, such as the rhymed couplet. However, settings of Lieder disrupt these norms for text depiction, primarily by phrase extension and contraction, to convey poetic tension and ambiguity (1996, p. 175).

For Neumann, when music is linked to words there is usually coordination between linguistic and musical structure. It is closest in recitative, in which musical declamation is fully adapted to the rhythms and inflections of the words. It is loose when, in arias or choruses, words or whole sentences are repeated and syllables extended in rich melismatic figurations. The closer the link, the more guidance from the text for the music's phrasing. In "closed" pieces or movements, such as arias, the link can be close if the words are mostly set syllabically (Neumann, 1993, p. 260).

Song-texts for phrase annotation

Many computationally based studies (especially in Natural Language Processing) use annotated databases to learn regularities and test models. Such databases could be useful for studies of musical phrasing. Although some exist (such as the Essen folk song collection),[3] there are currently no large databases of classical music annotated with phrase marks. As the text in songs is often considered to determine the musical structure, it may be that annotation using song-texts can be an (automatic) annotation tool for the identification of musical phrasing.

A study of Lieder

One of the aims of the introductory study (chapter 2) based on two Lieder by Schubert, is to investigate to what extent there is a direct link between the text structure of the poem and the phrase structure identified by musicians. It further aims to investigate of the degree and type of similarity or difference between responses, and how musicians treat the phrasing of the accompaniment and vocal parts and the relation between listeners' responses and performance and musical features.

A score annotation approach is used, limiting the investigation to 'musicians' (those who can play or sing from a musical score).[4] This introductory study also records the musicians' definitions of the term 'musical phrase' and their reasons for their phrase identifications. The musical phrasing of these pieces is further explored through an investigation of performance contours of publicly available recordings and a musical analysis and identification of the musical features. The distribution of the musical features is then compared with the phrases identified by the musicians and, in turn, these are compared with the structure of the text.

1.2.3 Listeners' phrasing of case-study pieces

The next group of studies investigates listeners' responses to different renditions of six case-study pieces. The analysis of pieces and the different responses forms a major part of this work in part to allow for direct comparison between different perspectives on the same music (chapters 3-8 and 10-11).

The theoretical and methodological aspects of the study of listeners' phrase perception include: 1) investigation of the extent of consistency of phrase identification, of the definition of the phrase, and the reasons for phrase identification, both those given by listeners, and those identified through the analysis of their results, and 2) methodological decisions concerning how the questions are to be posed and explored, and the type of music used: folk and/or classical, vocal and/or instrumental, polyphonic and/or monophonic, MIDI and/or performed.

The discussions of previous explorations and approaches to the term musical phrase in various disciplines and perspectives give rise to varied terminology and

---

[3] Initiated by Schaffrath and available at http://www.esac-data.org

[4] Responses of a population with more diverse musical experiences are discussed in chapter 3.

theories whose application is sometimes associated with different musical characteristics. Furthermore, the view that while performers intuitively understand what a phrase is, a precise and comprehensive definition is almost impossible is reminiscent of many. In general, a musical phrase is analogous to a sentence of prose or a line of poetry; all are more or less complete ideas that come to some sort of pause or closure and in music, such pauses are created by cadences Stein and Spillman, 1996, p. 174, see also Neumann, 1993, p. 260, cadences are discussed in section 1.2.10).

Previous studies investigated a relatively small number of musical elements (chapter 8). Here a relatively large number of musical elements are investigated together (chapters 10-11), including those used in psychological studies and those based on theories of music analysis (chapters 8-9).

Most studies that model phrase perception seem to assume that all listeners agree on phrase boundary positions (such as, Ferrand 2002; 2003, Cambouropoulos 2001; 2003) or model only a subset of the population (Lerdahl and Jackendoff 1987). Some even say that most ambiguities are present only in theory and that listeners can identify the 'correct' structure (Bod, 2001). Few studies have investigated consistency of phrase identification by listeners with different levels of musical training, and familiarisation with the piece used in the experiment. According to Deliège, musical education and degree of familiarisation with the piece seem to have a marginal effect on the results of the segmentation tasks (Deliège, 1998, p. 83, all these studies are discussed in chapter 8). Schaefer et al. (2004) on the other hand, find an effect of musical training. This leaves the aspects of musical training and familiarisation unresolved and these are therefore included in the current study (chapter 3).

No previous phrasing experiments explicitly elicit the listeners' verbal definition of a phrase. For a complete investigation of the term and the listeners' responses, however, it is necessary to investigate listeners' 'conscious' definition of the term, and to explore what possible 'template' they may be using. These can then be compared to those in the literature and their musical responses.

The music

Musical renditions

Listeners' phrasing experiments began in the late 1980s, concentrating on monophonic music, MIDI and performed renditions, of both short and longer pieces, and tested variables such as the effects of training and familiarisation. These studies provide partial information about phrase perception by listeners. Therefore, a broader study is carried out here (chapters 2, 3, 6, 7 and 10).

Different methods of exploring listeners' phrase (or segmentation) perception have been used. For example, Palmer and Krumhansl used preset segments and asked listeners to rate the degree of phrase completeness (Palmer and Krumhansl, 1987; 1987b), folk songs (Schaefer et al., 2004), and pieces from the musical canon (Deliège, 1987; 1998; Ferrand et al., 2002; 2003) are used for which listeners were asked to identify phrases after some listenings (see chapter 8). In this study, listeners (with different levels of musical training) are asked to identify phrases in a number of pieces from the musical canon (chapter 3).

Vocal and instrumental

Some modelling studies (Bod, 2001; Temperley, 2001), and parts of experimental ones (including Cambouropoulos, 2001 and Ferrand et al, 2002, 2003) are on folk songs (i.e. vocal music). Other studies carried out by music psychologists are on instrumental music (such as Deliège, 1998; Palmer and Krumhansl, 1987). This study investigates responses to vocal and instrumental music.

Monophonic and polyphonic music

The term phrase 'carries a melodic connotation, insofar as the term "phrasing" is usually applied to the subdivision of a melodic line. As a formal unit, however, it must be considered in its polyphonic entirety' (Macy).

The view that phrasing applies both to the melodic aspect of music and to the whole polyphonic texture, however complicated that makes its identification, is shared by most music-theoretical approaches, including those of Koch (1787, 1983), Rothstein (1989), and Keller (1965) who specifies both the complicated nature of polyphonic music and the importance of considering all the voices. However, most of the psychological and computational studies use either monophonic folk-songs, such as from the Essen Folk song collection (Huron, 1996, Bod, 2001, Temperley, 2001, Ferrand et al., 2002, Schaefer et al., 2004, and some examples in Ferrand 2003), monophonic classical music (Deliège, 1998 and the rest of the examples in Ferrand et al., 2003), or music that is made monophonic for the study by taking the melody line only (such as in Lerdahl and Jackendoff, 1987, p. 37, Palmer and Krumhansl, 1987, Ferrand et al., 2002 and Cambouropoulos, 2001; 2003).

There are clear reasons for using folk music. It has been developed and sung by people with no need of special musical training and was not learnt from notation. Moreover, in this (and any other vocal music with words) there are texts that can provide other, non-musical, yet integral cues and thus possibly provide more phrasing information, though, the texts are not used in the discussion or analysis of this music in any of the studies (such as Bod, 2001 and Cambouropoulos, 2001; 2003).

However, by using only monophonic music, much information is not studied. For example, though western tonal monophonic music contains implied harmony that is 'heard' by performers and listeners, as it is not explicitly represented 'in the score', and as so much else is, harmony is usually excluded from the main discussion and explanation (though not always, see Lerdahl and Jackendoff, 1987).

The study of monophonic music also avoids the question of phrase identification in different simultaneous voices. This identification is problematic partly because it is unclear how the different voices contribute to phrase perception. For example, are individual voices treated separately and therefore phrased separately? Does a single voice always dominate? Is there a combination of the two depending on the musical context and listener? All the theorists quoted above mention this problem. Temperley (2001) even describes the perception of phrasing in a polyphonic piece, but does not include it in his model. From a computational perspective, the problems of analysing polyphonic music have been approached in other domains (Longuet-Higgins and Lisle, 1989, pp. 21-22; Temperley, 2001) but not phrasing. A more comprehensive approach would provide a more representative sample and responses to help investigate this question and is therefore employed here (chapters 3 and 10).

MIDI and performed renditions

Different types of renditions have been used in listeners' phrase studies: MIDI (Palmer and Krumhansl, 1987) and recorded performances (Deliège, 1998). These two types of renditions, however, have not been used in the same study, so it has not been possible to compare responses directly. In this study, both types of renditions of the same pieces are used (chapter 3).

Having collected listeners' responses to monophonic and polyphonic western classical music in MIDI and performance renditions, two different analytical approaches were employed.

1.2.4 Classification agreement analysis

The listeners' responses are analysed in two ways. The first is based on a combination of music-analytical ideas and numerical comparisons (chapters 3 and 10). The second is based on a method previously developed for the analysis of responses to linguistic stimuli and uses the Kappa statistic (Krippendorff, 1980), quantifying the degree of consistency within and between listeners (chapter 4). This method has not been used previously in musical contexts and is here adapted and developed for it. As the same method is used as in the linguistic domain, this may enable comparison between responses in these related domains. The methodology and computational implementation were developed with Beata Beigman Klebanov at the Hebrew University, Jerusalem.

1.2.5 Click detection

Another approach to investigating listeners' phrase perception is to focus on the precise location of the phrase boundary by applying an indirect method. The current method is based on a theory that we respond to extraneous noise more quickly when the 'cognitive load' is smallest (i.e. at phrase boundaries) and that we 'wait' for low cognitive load 'to deal with' the additional information. The theory and experimental approaches were developed in studies of musical and linguistic processing, primarily for the exploration of peoples' responses to segment boundaries. The common element to these studies is the exploration of responses to a super-posed click. In all of the studies, the responses to the clicks are expected to be different depending on the position of the click relative to a syntactic boundary (clause, sentence, or phrase). In this study, this method is applied to the case-study pieces to explore both its usefulness for this purpose and the responses to different types of phrase "boundaries" in these pieces (chapter 5).

1.2.6 Musicians' annotation of the case-study pieces

Most of the experimental studies of phrasing have used different types of visual representation and listeners with different musical experience (Deliège, 1998; Palmer and Krumhansl, 1987). Some computational studies (such as, Bod, 2001; Temperley, 2001) use data-bases annotated by a single musician or possibly a small group of musicians arriving at a consensus. It is not clear how this consensus was reached or how much variability there was in phrase identification. Lerdahl and Jackendoff, study grouping as identified by the authors themselves, leading to a model of perception by the 'experienced listener' (1987, p. 1).

The current study of musicians' annotations (marking phrasing as if in preparation for performance) is carried out on some of the case-study pieces of the listening study (chapter 6). It investigates this assumption of consensus in a situation similar to that of the above studies by questioning only musicians and asks for responses based on the "score" (as in chapter 2). The study investigates the agreement between responses and compares between these score-based responses and the listeners' responses to, and performance contours of, the same pieces (chapter 10). The performance contours are presented in chapter 7.

1.2.7 Performers' phrasing study

Performers rarely write about the process of preparing a performance and even less about phrasing explicitly, though they use the term frequently. The few published writings include some repeating themes such as that the music is 'divided' and how the phrases relate to each other (such as in Kirkpatrick, 1984). Indeed, for some, in tonal music, phrasing and articulation are two of the chief elements for which the performer bears the most direct responsibility (Chew). Macpherson, providing guidelines for performance, emphasizes the importance of phrasing. He highlights cadence bars, and that '[s]imilar (melodic or rhythmic) patterns imply uniformity in phrasing, contrasts of melodic or harmonic ideas in quality or quantity of tone, or pitch should be taken as indications of breaks or divisions in the rhythmic groups, and groups of quick notes frequently end in a longer one (accented or not) and should be phrased so as to include that note'

(Macpherson, 1912, p. 19). Unlike other theorists, Macpherson makes a direct connection between slurs and staccato marks, and phrases determined by elements such as cadences.

Pieces may have different structural interpretations, while expression in performance may limit the extent of this ambiguity (Clarke, 1988, p. 15). This and other studies suggest that phrase structure is reflected and clarified in performance and is one of the central musical elements that contribute to the way a piece is performed and therefore perceived: 'phrasing tends to dominate performance expression' (Friberg and Battel, 2002, p. 207).

Previous studies have identified tempo and intensity changes as being important for identifying phrase boundaries, particularly ends (Gabrielsson, 1987; Hartmann, 1932; Povel, 1977; Repp, 1990). Here, recorded performances of the case-study pieces are investigated using similar methods as the above studies, enabling a comparison between observed performance characteristics and responses for the same pieces (chapter 7). Having collected all the data for this study, the specific examples of the music psychological literature are returned to in order analyse to what extent the phrase characteristics that have been previously identified explains the responses collected. In more general terms, this comparison enables the identification of what might be missing from current models and definitions (chapter 8). At their basis, many of the approaches take fundamental psychological theories, which are here discussed together with other more general ones.

1.2.8 Previous psychological approaches

Many of the studies that investigate phrasing base their approaches on Gestalt principles, seven studies reported in the literature will be discussed in chapter 8. Here the application of Gestalt psychology in a musical context, and other psychological theories are described.

1.2.8.1 Gestalt psychology

Gestalt psychology, developed with the aim of being applicable to perception in any domain, forms a basis of theoretical works that have been considered fundamental in the study of phrasing (including, Lerdahl and Jackendoff, 1987 and Deliège, 1987; 1998) and other aspects of music perception (Deutsch, 1999; Handel, 1989). The approach has provided the most complete description of factors influencing grouping and segmentation of music. Though these principles do not constitute an explanation of the segmentation of music, they provide testable generalisations about it.

> 'Man könnte das Grundproblem der Gestalttheorie etwa so zu formulieren suchen: Es gibt Zusammenhänge, bei denen nicht, was im Ganzen geschieht, sich daraus herleitet, wie die einzelnen Stücke sind und sich zusammensetzen, sondern umgekehrt, wo - im prägnanten Fall - sich das,

was an einem Teil dieses Ganzen geschieht, bestimmt von inneren Strukturgesetzen dieses seines Ganzen'[5] (Wertheimer, 1924).

This is the "fundamental formula" of Gestalt theory. Two features are explicit: 1) The control of the whole over the part, and 2) what happens in a part is determined by the structural laws of its whole. So, 'Gestalt' refers to the structural laws of the 'whole'. In particular, Gestalt psychologists investigated how the perception of a whole could be greater than the sum of the percepts of its parts. For them *form* is the primitive unit of perception (Köhler and Wallach, 1944).

The Gestalt approach was the first systematic attempt to study perceptual segregation and organisation to which it gives rise (Eysenck and Keane, 1995, p. 33). From the early discussions, questions addressed mainly through visual perception are almost always accompanied by what Wertheimer considered to be the parallel in musical perception. Instead of hearing a melody as a sum of individual tones that constitute the primary foundation of the experience '…what I hear of each individual note, what I experience at each place in the melody is a … part which is itself determined by the character of the whole…. The flesh and blood of a tone depends from the start upon its role in the melody: a b as leading tone to c is something radically different from the b as tonic. It belongs to the flesh and blood of the things given in experience [Gegebenheiten], how, in what role, in what function they are in their whole' (Wertheimer, 1924, 1938).[6] However, in 'a Beethoven symphony…it would be possible to select one part of the whole and work from that towards an idea of the structural principle motivating and determining the whole. Here the fundamental laws are not those of fortuitous pieces, but concern the very character of the event' (Wertheimer, 1924, 1938). This inconsistency indicates difficulties for the application of this approach to western classical music. Phrasing, as such, is not mentioned by Wertheimer, but his questions, theory and rules lead to some of the 'psychological' characteristics that have been considered important in phrase perception.

1.2.8.2 Gestalt Factors

Through a discussion of a series of discontinuous (mainly visual) constellations, Wertheimer presents 'factors' – the definite principles governing perceived

---

[5] 'There are wholes, the behaviour of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole. It is the hope of Gestalt theory to determine the nature of such wholes. With a formula such as this one might close, for Gestalt theory is neither more nor less than this. It is not interested in puzzling out philosophic questions which such a formula might suggest' (trans Ellis, 1938, p. 1).

[6] In transposition, the sum of the elements is different, yet the melody is the same; one is often not aware of the transposition: 'There must be something more than the sum of six tones. viz. a seventh something, which is the form-quality, the Gestaltqualität, of the original six. It is this seventh factor or element which enabled you to recognise the melody despite its transposition' (Wertheimer, 1924, 1938).

arrangements and divisions (a more detailed presentation of the principles is given in appendix 1). The fundamental principle is *Prägnanz*: 'Of several geometrically possible organisations that one will actually occur which possesses the best, simplest and most stable shape' (Koffka, 1935) and most of the other principles can be subsumed under it (Eysenck and Keane, 1995, p. 33).

1. The Factor of Proximity
2. The Factor of Similarity
3. The Factor of Uniform Density (or of "Common Fate" – the shift)
4. The Factor of the Objective Set [*Einstellung*]
5. The Factor of Direction
6. The Factor of Good Continuation/Inner Coherence/Good Gestalt
7. The Factor of Closure (Law of enclosedness)
8. The Factor of Past Experience or Habit

Together the factors may co-operate or be set in opposition and it is possible to test their relative strengths (Wertheimer, 1924). In the musical context, this has been questioned (including, Handel, 1989; Howell et al., 1991, p. 20). Quantity does not affect the ease with which we unite elements into groups. Only 'unnatural, artificial arrangements' become more difficult with more elements (Wertheimer, 1924). Insightful learning (finding the principle or pattern) is more durable than rote rehearsal, demonstrating the crucial rôle of the subject as an active processor and organiser of material rather than a passive recipient (Baddeley, 1990, p. 129).

1.2.8.3 Music and Gestalt psychology

The main grouping factors that have been discussed in the musical context are 1, 2, 6, and 7 (above) and have been examined most systematically with regard to separation of musical lines in pseudo-polyphony but may operate on any of the perceptible attributes of musical elements and subunits. Theoretical, experimental and computational approaches have aimed to define more clearly the musical meaning and use of some of the terms (e.g. Cambouropoulos, 2001).

Tenney and Polansky (1980) developed a formal computational model to test the Gestalt principles of proximity and similarity[7] for identifying segments of monophonic test pieces. They looked at pitch intervals, initial intensity, final intensity, duration and rest-duration (1980, p. 218). All of these should be weighted, though, they explain, no clear principle has been discovered for determining what the weights should be, and these may be piece-dependent. In spite of the limitations (1980, p. 217), the correspondence of these results to segmentations arrived at by other (music-analytic) suggests that the 'fundamental hypothesis of temporal gestalt perception' is at least a plausible formulation of a

[7] Similarity can be considered to include proximity as a special case (Tenney and Polansky, 1980, p. 211).

principle of musical perception (1980, p. 236). The segments identified are very short in comparison with most segments referred to as phrases by theorists and may be subphrases at best, so the method cannot be transferred to phrasing directly. This approach, however, shows that a computational model can test some of the Gestalt principles in the 'segmentation' of music and that the output is quite similar to results of some musical analysis.

Lerdahl and Jackendoff's (1987) model also used musical interpretations of some Gestalt factors in the form of a rule base for grouping structure in music (including phrasing). This account is the most systematic theoretical approach of this kind and has been empirically investigated (Deliège, 1987). Their work and that of others, which in many ways follows from it, are discussed in chapter 8.

There have been criticisms of Gestalt theory ranging from the basic starting point, to the principles, to the physical manifestation in the brain (Hochberg, 1998, p. 288). However, generative linguistics has helped rekindle interest in mentalistic theories and Gestalt theory though the physiological reductions proposed are too crude for the fine observations they are meant to explain (Lerdahl and Jackendoff, 1987).

1.2.8.4 Gestalt factors and the study of phrasing

Gestalt principles have been used in phrase perception studies though no author suggests that these alone could explain all the characteristics observed. There are several elements of the theory that seem suitable for the study of musical phrasing: The premise that the whole is 'greater than the sum of its parts', that it controls the perception of the parts more or as much as the parts do themselves, and that the same parts within different wholes can be perceived differently. The aim of the theory, to provide a basis for scientific exploration and more specifically to be applicable to various types of perception (such as vision and language), is appealing.

However, there are problems specific to phrase perception. The Gestalt theorists and some of those who applied the Gestalt factors to music stated that the principles should be applicable at all levels. Though studies have concentrated on segmentation of many levels, they have not always specifically referred to the phrase level. Bod (2002) criticises the application of the Gestalt factors for phrasing explaining that the Gestalt-based, parallelism-based (as well as harmony-based models) are inadequate for some gradient phenomena (chapter 8.5). Moreover, even when looking at the phrase level, identifying the phrase segments is only part of the current aim.

1.2.8.5 Phrasing, memory and the psychological present

Snyder distinguishes between melodic and rhythmic groupings, and phrases; closure is usually established more completely at the phrase level, though the distinction is not absolute. Phrase boundaries are usually reinforced by changes in more than one parameter such as the combination of a pause following a relatively long note. For him, phrases are the largest unit of musical material that can be accommodated by short-term memory and therefore the largest units of musical experience that can be completely integrated in the present. Therefore, themes designed to function as a unit to be transformed are usually no longer than a phrase. Musical phrases are often linked to some variable of human physiology, such as how many events can be produced in a single breath on a wind instrument, or in a single bowing movement of a stringed instrument. Thus, a musical structural rhythm is superimposed on a basic human physiological one. These are single coherent physical gestures, and the action components of physical gestures are also 'chunked' (Snyder, 2000, pp. 37-39).

Above a threshold, our memory of phrases is more schematic (Snyder, 2000, p. 39). The time limit of this short-term memory can vary with the amount of information it is processing, occasionally reaching as long as 10-12 sec although the average is 3-5 sec, or approximately seven chunks (Miller, 1956). Listeners' short-term memory limitations affect the extent to which their representation decisions could be postponed pending additional input (Snyder, 2000, p. 50). Failure to construct a representation within the span may lead to the loss of the information. Representation decisions must be made with almost no delay based on a limited number of events, but the input available to the listener at the time does not lend itself to a unique representation; representation decisions must be made when the musical grammar is still unable to suggest a solution (Berent and Perfetti, 1993, p. 204).

The psychological present is 'that part of our ongoing experience currently accessible to consciousness' (Dowling and Harwood, 1986, p. 179, see also James 1890, p. 609). The 'part' varies with attention and characteristics of the material and can be manipulated by composers and performers (Dowling and Harwood, 1986, p. 179-181). Fraisse's studies indicate that the psychological present typically extends less than 5 sec (1978; 1982). Listeners accurately reproduced simple sound sequences 3 or 4 sec long and "chunked" longer sequences of clicks; they are able to accurately perceive sequences as long as 25 clicks by chunking them into five groups (see also James, 1890). James cited Wundt and Dietze stating that the extreme upper limit of accurate recognition of a click series is about 40 over a period of 12 sec, which could be chunked into either five groups of 8 or eight groups of 5 (Dowling and Harwood, 1986, p. 180) agreeing with Miller's (1956) estimate.

Phrase lengths in songs and poetry have been related to the psychological present citing ranges of between 2 and 5 sec (Fraisse, 1982). For example, Dowling (1984) found that the average time per phrase in songs sung by children was 5.50 sec at around 18 months and 4.47 sec at around 3 years. The decrease is probably due to minimizing pauses between phrases as part of a general tightening of song structure. Singers accompanying a children's songbook sang at a rapid rate without pauses between phrases, giving a mean phrase duration of 2.52 sec. In contrast, for Joni Mitchell (on *Miles of Aisles*), the average phrase length was 4.32 sec (range 2.54-5.36).[8] There were, like in the children's performances, brief pauses between phrases. Comparison of Mitchell's performances with that of Judy Collins of the same song showed that they were very similar, while Mitchell's were more different on her own later recordings (Dowling and Harwood, 1986, pp. 180-181).

The psychological present, therefore, is thought to be in the range of 2–5 sec, occasionally reaching 10 or 12 sec, and does not appear to change radically from early childhood.

These fundamental psychological observations are combined with music -theoretic and -psychological ones for the study of musical phrasing. Following the analysis of specific music-psychological approaches, two fundamental music-theoretic ones are discussed, preceded by a brief discussion of music analysis to understand its aims and general relevance to the understanding of the perception of phrases (chapter 9).

1.2.9 Musical analysis and phrasing

Music analysis includes the interpretation of structures in music, their resolution into simpler constituents, and the investigation of the relevant functions of those elements, embraces a large number of diverse activities, and different methodologies and theories, have been developed (Bent and Pople). Theories and approaches (particularly those of Koch, 1787, 1983 and Rothstein, 1989) are discussed in chapters 9. This prepares for their use in the analysis of the case-study pieces in chapter 10. Based on the previous research and the analysis, a number of musical features are identified, the frequency and importance of which are analysed in more detail in comparison with listeners' responses in chapter 11.

Here, a short summary is given of some of those musical elements that form much of the basis for the musical discussion in this study (chapters 9-13). Most theorists begin by mentioning the broad categories of characteristics and elements that they consider to be most important for phrasing and phrase identification and then give more or less detail about these elements.

1.2.9.1 Structure, expectation, emotion

Even when playing from a score, unless the music is already known well, uncertainty surrounds the anticipation of how the music may develop, and the relationship between current and future events. The future course of the music is not completely unknown, as listeners and performers continually make projections on the basis of acquired stylistic knowledge (Meyer, 1956). However, the music

---

[8] Long instrumental interludes were not included. The comparison could have provided important information.

seldom matches these projections as 1) performers and listeners may envisage a number of different possible continuations, 2) stylistic knowledge is never perfect and complete, and 3) it is the nature of music to depart from stylistic norms in the interest of creativity (Clarke, 1988). Implications or expectations and their realisations of different parameters have been investigated in terms of pitch,[9] harmony, and expectation related tensions.[10]

Expectations and the ways in which they do or do not resolve has been seen as important in some approaches to music and emotion, for the listener and performer (for example, Meyer, 1956, Clarke, 1988 and Friberg and Battel, 2002, p. 199). However, the relation between structure and emotion has been hotly debated (see Juslin and Sloboda, 2001). Many of the features mentioned in these studies are considered here.

1.2.9.2 Ambiguity and irregularities

In general, ambiguity refers to information that has already been heard, while uncertainty refers to the idea that the listener does not know what will follow. The term ambiguous is often used in studies about phrasing but seems to have different connotations. For some (such as Bod, 2001), while the listener's intuitions are clear as to where phrase boundaries should fall, ambiguity is a computational problem to be solved. For others (such as, Temperley, 2001), the difficulty is for the listeners; our intuitions as to if and where phrase boundaries occur can be vague (see also, Friberg and Battel, 2002, pp. 205-6).

Certain phrases are described as irregular. For example: 1. Connecting figures may mask an otherwise obvious cadence (Lampl, 1996, pp. 104-5).[11] 2. Individual phrases, which initially seem to conform to a 'regular' four-bar pattern, turn out to be shortened (elision) or lengthened (extension). 3. Adjoining phrases may be linked through 'dovetailing' in which the last bar of one phrase (which is elided) coincides with the first bar of the following phrase (Lampl, 1996, pp. 104-5).[12] 4. Phrases can overlap, as in fugues and other imitative counterpoint (Lampl, 1996, pp. 104-5).[13]

---

[9] Narmour (1990; 1992), Schellenberg (1997), Povel and Jansen (2002; 2004).

[10] Jansen (2004), Krumhansl and Kessler (1982), Krumhansl (2000), Bigand and Parncutt (1999), Schmuckler (1989).

[11] See also phrase-end concealment: the boundary between two phrases is deliberately effaced or bridged over, such as in Mozart Sonata K. 333 (Keller, 1965, p. 21).

[12] See also phrase elision: the change is so abrupt that the final tone of the first phrase is not heard (Keller, 1965, p. 24).

[13] See also Phrase linkage: the first phrase ends with the first note of the next. This is almost the rule in expositions and fugue stretti. The newly entering voice starts on the final note of the first, not permitting the motion to come to rest. This impression of a new beginning can also arise through a sudden diminuendo, usually requiring the presence of two or more voices (Keller, 1965, pp. 23-5)

These phrase irregularities all seem to be set deviations from "standard" phrases indicating that even when phrases are irregular they can still be related to an expected, "standard" phrase structure.

1.2.9.3 Hierarchy

Hierarchy is one of the aspects of musical structure most often discussed. Structures are represented as organized in a series of levels related by reduction or elaboration (Clarke, 1988). Though evidence for completely unified structural knowledge is problematic, aspects entailed by hierarchical structure seem to be used in music perception (Krumhansl, 1983).

In general terms, for example, each of the large sections in a movement may be seen as unified at a high level and, at the same time, each containing a multi-levelled branching structure illustrated by trees, relating events at lower hierarchical levels (Clarke, 1988). For Clarke, if a performer has memorized a piece, at any time, part of the hierarchical structure is active, the rest being in a latent state, or active in broad outline. For instance, in the middle of a deeply embedded musical phrase, only a region of low-level generative connections might be active, since there is little need for a performer to have access to high-level structural information. At a phrase boundary, however, it may be important to know how the previous and subsequent phrases are related to one another and to the overall structure of the piece (1988, p. 4).

Phrase structure is often, assumed to be hierarchical (for example, Friberg and Battel, 2002; Neumann, 1993).[14] 'Most tonal music has a hierarchical phrase structure, sometimes simply called *grouping*. The slowest level is the entire piece, which is then divided and subdivided into sections, phrases, subphrases and melodic groups. Superimposed upon this is usually a *metrical hierarchy*: the beat or *tactus* is grouped, usually in groups of two or three, into measures and groups of measures. The beat can be divided into subbeats' (Friberg and Battel, 2002, p. 201). The fastest level in the phrase hierarchy consists of small melodic units of a few notes each. Grouping (i.e. segmentation) at this level tends to be quite ambiguous, often with several possible interpretations. So communication of this structure can be subject to more individual interpretation than longer phrases. Friberg and Battel discuss the Mozart A-major sonata; many performers choose

[14] As illustrated by some definitions such as Neumann's: There are two aspects of melodic (phrasing) structure. 1. 'Anatomy' – segmentation of melody in distinct units; larger ones may be composed of smaller ones, which in turn can contain subunits in an often complex hierarchic pattern and 2. 'Physiology' – the way in which single pitches of a melody, meaningless by themselves, generate, through their interaction, an energy flow between levels of lower and higher intensity e.g. from impulse, over climax, to repose. Difficulties for anatomic aspects of phrasing vary with the style and texture of the music involved (see above Historical Development: by era or by composer). Difficulties for physiological aspects, the energy flow within a melodic sequence, are considerable regardless of style (Neumann, 1993).

the first five notes as a group while others choose the first four notes, giving the second group an upbeat. This ambiguity may be due to contradictory perceptual cues from different aspects of the musical structure, such as the melodic contour or the meter, and, according to Friberg and Battel, can be resolved in performance by inserting a micropause between the last tone of one phrase and the first of the next, which both interrupts the sound and delays the onset of the following tone (2002, pp. 205-6, Friberg et al. 1998, and Clarke, 1988).

Some researchers use the term "hierarchy" more formally than others. Lerdahl and Jackendoff, explain that hierarchical structures in general organisations

> 'composed of discrete elements or regions related in such a way that one element or region subsumes or contains other elements or regions. A subsumed or contained element or region can be said to be *subordinate* to the element that subsumes or contains it...In principle this process of subordination (or domination) can continue indefinitely. Thus all elements or regions in a hierarchy except those at the very top and bottom of the structure are subordinate in one direction and dominating in the other. Elements or regions that are about equally subordinate within the entire hierarchy can be thought of as being at a particular hierarchical *level*. A particular level can be spoken of as *small-scale* or *large-scale*, depending on the size of its constituent elements or regions.
> In a strictly hierarchical organization, a dominating region contains subordinate regions but cannot partially overlap with those regions' (p. 13).

They explain that grouping structure is 'recursive' i.e. it can be elaborated indefinitely by the same rules, and that nonadjacent units cannot be grouped together (p. 16). They further explain that grouping is not '*strictly* hierarchical' as evidenced by overlapping elided phrases. However, 'the conditions under which overlaps and elisions are perceived are highly constrained' (p. 14). Moreover, a 'recursive' nature, particularly the indefinite elaboration by the same rules, does not seem applicable to phrasing and this is one of the aspects that distinguish the phrase from the 'group'. From their discussion of hierarchy there is an implication that phrases form one level of this grouping hierarchy but there is no direct expression of this. Hierarchy can also refer to inclusion of smaller structures (such as phrases) within larger ones.

As there are phrase structures that do not follow the strict hierarchical rules it seems that rather than identifying a complete hierarchical structure in phrasing, only certain aspects of such a definition of hierarchy may play a rôle in phrase structure and perception (chapter 13).

1.2.10 Musical features

From the general discussions above, it becomes increasingly clear that for a better understanding of the perception of the phrase, it is essential to study the individual rôle of different musical features in phrases. They are briefly reviewed here in the context of general music theory, music analysis, music psychology and computational musicology and are discussed in chapters 8 and 9-14.

1.2.10.1 Pitch jumps, rests, long notes and pauses

Gaps in the music are often seen as boundary markers, especially from the perspective of the Gestalt principle of proximity. A relatively large pitch interval is seen as equivalent to a gap in a visual stimulus. This element tends to dominate the psychological and computational literature (chapter 8) and does not feature as strongly in the music-theoretic literature.

Rests are discussed in most approaches. For some they are equivalent to long notes, the concentration being on the gap between note onsets, while others distinguish between the two, and for some they are very important (for example, Lampl, 1996, p. 102-4). Music psychologists and computational approaches that use either, usually relate them, like pitch jumps, to the Gestalt principle of proximity (chapter 8).

In some cases, however, long notes may be delaying a resolution (through for example, suspensions) and are therefore not simply gap features. Another warning comes from Rothstein who explains that Chopin frequently uses the rhythm  as a sprightlier variant of  without implication of a break in the phrase or sub-phrase; they are elements of articulation, not phrase structure indicators (1988, p. 123). Most psychological approaches only consider rests or 'long notes' that are longer than the preceding and following ones as gaps which could be boundary indicators (such as Lerdahl and Jackendoff, 1987, see appendix 8.2, and Temperley, 2001). Many early, and some more recent, texts emphasise the use of pauses for phrase end identification (including, Lampl, 1996, pp. 102-4).

1.2.10.2 Slurs

Lampl includes slurs as equivalent to punctuation in language though, for him, they denote 'groups' rather than 'phrases', indicating that slurs are not as clear as other elements (Lampl, 1996, see also Schenker, 1925, 1994 and Rothstein, 1988).[15]

1.2.10.3 Changes

Change is a Gestalt principle used in phrase theories (chapter 8) and is also used in music-theoretic approaches in terms of, for example, changes in range and

[15] For example, Chopin's slurs, according to Rothstein, are an analytical minefield. 'Chopin's practice alone should be proof enough that legato articulation and phrase structure ('phrasing') are inherently different aspects of music, related only in so far as the former may be used to delineate the latter' (Rothstein, 1988).

instrumentation (Lampl, 1996, p. 102-4) and can be in, for example, texture, motive, and harmony, some of which are rarely investigated in the context of phrasing.

1.2.10.4 Repetition

Repetition often features among the elements contributing to phrases (for example, Lampl, 1996, pp. 102-4). Repeats (or similarity) are included in many of the Gestalt-based approaches. Identification of exact repetition of more than a few notes is relatively straightforward and considered important by many authors for the identification of phrases. Some sequences of notes that are repeated in the same piece but are modified, are also recognised as repetitions. Methods of identification of these has been approached many times and in different fields from the perceptual 'segmentation' (Cambouropoulos, 2001) to 'metre' (Steedman, 1977), to practical applications, such as identifying themes for music retrieval systems (Meredith et al., 2001 and Damiani et al., 2003), using computational techniques and some listeners' responses studies.

1.2.10. Goal directed and harmonic motion, expectation, breath, and the cadence[16]

1.2.10.5.1 Breath, expectation and the goal

Many authors discuss phrasing in the context of goal (or climax) directed motion or progression (such as Barra, 1983; such as Macpherson, 1912),[17] sometimes with the idea of breath. Sessions (a composer) adds another physical dimension of holding on or rather, not letting go (1950, p. 13). The idea of goal directed motion, is sometimes extended and described as a curve, the focal point of which may occur in different positions producing impulses of different accents (Barra, 1983, p. 38).[18] Westergaard, in a textbook on tonal theory, also emphasises the

---

[16] Cadence, from *caderer* (latin) meaning, to fall or land.

[17] 'Everything in music must be considered in the light of progression, or movement towards some more or less clearly defined destination' through the gradual but inevitable working-up of some extended passage to a strong emotional climax, or of a 'trend' of some figure towards the point where it finds its own completion. In a musical period, such as is understood by a phrase or a sentence, the rhythmic climax is at the cadence at its end, and all that precedes this leads to it (Machperson, 1912). In phrasing 'the most important quality of any musical action is its sense of forward momentum or thrust. Controlling this momentum – nurturing, reinforcing, guiding, shaping and ultimately resolving it – is one of the most crucial aspects of any musical performance' (Barra, 1983, p. 19).

[18] For example, the phrase 'is based upon the principle of the dynamic curve. Typically, tonal actions begin with an anacrusis, or growth phrase of increasing energy, reach a focal point of highest intensity, then end with a concluding phrase, a release or relaxation. The focal point may occur near the beginning or the end of the phrase, to produce beginning-accented or end-accented tonal impulses. These internal phrases normally for several higher-level cycles of motion, each with its own pattern of development' (Barra, 1983, p.

movement towards a goal and the sense of completion and describes the temporal aspect of expectation (1975, p. 311).

One of the most in-depth discussions and descriptions of the phrase is given by Rothstein: 'a phrase should be understood as, among other things, a directed motion in time from one tonal entity to another; these entities may be harmonies, melodic tones (in any voice or voices), or some combination of the two. *If there is no tonal motion, there is no phrase*' (Rothstein, 1989, p. 5, discussed in more detail in chapter 9).

1.2.10.5.2 Harmony and cadence

The construction and effect of harmonic structures and their relationships is one of the fundamental aspects of western classical music.[19] Phrase and harmonic structures (particularly phrase ends, goals and cadences) are often described hand in hand, sometimes defining one by the other, especially in the music-theoretic literature.[20] However, there are few such cases in the music-psychological/computational theories.

Harmony, cadences and the phrase, interlinked
An important factor 'in the interpretation of phrases is the harmonic structure of a progression. Harmony provides a tonal skeleton against which melodies develop, and it also creates a pattern of motion that contains its own sequence of thrusts and resolutions' (Barra, 1983, p. 51).
Closure often involves an acceleration of harmonic rhythm at the approach to the cadence. The increase in harmonic activity emphasizes the close of a phrase or piece. A large tonic reprise or final tonic cadence is sometimes prepared by a dominant pedal, a slowing down of harmonic rhythm that emphasizes the imminent resolution of V to I (Stein and Spillman, 1996, p. 173).
The cadence is one of Lampl's elements equivalent to punctuation in language, defining the boundaries of musical ideas. He cites the Harvard Dictionary of Music: 'Cadences which clearly end a phrase, "conveying the impression of a momentary or permanent conclusion" ' (Lampl, 1996, pp. 102-4).
Traditionally, the formal unit considered to be closed by a cadence is the *phrase*. Cadence and phrase are so intimately connected that the two terms are frequently defined in reference to each other, a cadence is a melodic-harmonic

21). In performance, the crucial position is just before the resolution, the point of highest tension, when a loss of intensity can have the most unfortunate effect (Barra, 1983, p. 38).

[19] Several perceptual models of harmonic structure have been proposed. Many take music theoretic or informatics approaches, including, Temperley (1997, 2001), Leman (1995), Parncutt (1989), and Povel (2002), see Jansen (2004).

[20] Some regard the close connection of phrase and cadence, and the repercussions of it, as problematic. For Caplin, cadence and phrase should be disengaged. *Cadence* can be viewed as a manifestation of formal funcationality, whereas *phrase* can be a functionally neutral term for grouping structure (embracing approximately four bars). It is then possible to describe which phrases have cadential closure and which do not (2004, p. 59).

ending of a phrase; a phrase is a formal unit ending with a cadence (Blombach, 1987, p. 226).
The phrase is a constant motion toward a goal – the cadence (Sessions, 1950, p. 12).
A phrase can be roughly characterised as the lowest level grouping which has a structural beginning, a middle and a structural ending (a cadence) (Lerdahl and Jackendoff, 1977, p. 123).
The conclusion to a phrase, movement or piece based on a recognisable melodic formula, harmonic progression or dissonance resolution; the formula on which such a conclusion is based' ('Cadence', in *Grove*)
The concept of phrase is most productively understood, both historically and theoretically, as admitting only two choices for its end-point: a half cadence or an authentic cadence (Darcy and Hepokoski, 1997, p. 123)

Harmonic progressions are often seen as leading from one place to another either by long term, journey-emphasising progressions: cycles of fifths, cycles of thirds, or short-term, end-emphasising progressions, such as cadences. There are several descriptions, suggested rules and theories concerning the structure of harmonic progressions and their perception. When harmonic progressions are discussed, this is rarely alongside some of the other features discussed here such as pitch intervals.

Like many music-theoretic terms (including 'phrase'), 'cadence' is commonly used but can refer to different musical phenomena and debate continues about what kind of harmonic progressions indicate what kind of phrase ends.[21]

The importance of cadences in western classical music stems from the need for a tonal centre (an anchor) that is established, elaborated, travelled away from, counteracted by a contrasting tonal centre and returned to. The structures occur on several levels between and within movements, sections, and phrases.

The harmonic characteristics of the phrase not only help define its boundaries and expectations thereof, but its internal structure, the relation between consecutive phrases, and the relative harmonic (and structural) importance of them in a movement (or section).

There are many different types of cadences and each has a number of possible chord combinations. The most straightforward, and usually strongest, type of cadence is the 'perfect' (or final or full) cadence:

[5 – 4 –] 3	– 2 –	1
[I/Iib/7th/Dim chord]	Va	Ia

If the I arrives on the first beat of a bar it is said to be stronger than one arriving on a weak beat. In either case, the phrase end coincides with this arrival. Arrival

[21] Of the vast literature on harmony and cadences, here the discussion concentrates on aspects related to phrase perception.

may occur at different times in different parts or may be extended after it has first been sounded. Many pieces from the styles discussed in this study end with such cadences. These progressions are often highlighted and exaggerated temporally (with, for example, extended pedals on the dominant and or tonic) and/or texturally. This combination may be seen as associated with, or representing the end. This association is explained in many composition manuals (such as, Koch, 1787, 1983), used and explained by analysts (Rothstein, 1989), and made into a rule in the work of music-psychologists (Lerdahl and Jackendoff 1987). This association with ending seems to be generalised for listeners (see, for example, probe-tone experiments, such as by, Krumhansl and Kessler, 1982, though, Thomson, 2001, pp. 131-3 questions the basis of these experiments) and seems to affect the way we perceive phrases in general. The cadence may not resolve to the tonic in the melody (or bass), the final chord may be replaced with the relative minor, or the progression may end on the dominant.

The cadence is treated here as a combination of vertical harmonic sonorities and is primarily defined in terms of its harmonic close. However, it is not possible, or useful to separate harmonic considerations from other musical features such as melodic (Caplin, 2004, p. 57), rhythmic, metric (Reimann, 1903, Meyer, 1973, Caplin 2004, p. 57), or those of length (Caplin, 2004, p. 57). Furthermore, in some cases, the harmony may be monotonic or too complex, in which case this feature would not be useful.

1.2.10.6 Voice-Leading

Voice-leading progressions may play a rôle in phrasing, particularly towards the ends of phrases.[22] Step-wise descents from the 8th, 5th or 3rd degrees of the scale, arriving on the 1st may be particularly important. They may arrive directly or be interrupted. Moreover, they may be explicit (as in a descending scale), or implicit in, and forming the basis of, a descending progression or sequence. The priority of pitch explains why the basic level of rhythmic analysis in Schenkerian contexts is usually the phrase, since the phrase is the smallest level of complete linear motion (Kramer 1988).

1.2.10.7 Melody

These ideas are related to that of pitch patterns throughout the phrase. 'One of the most important factors in determining the dynamic form of a phrase is the contour of the melodic line' (Barra, 1983, p. 47) which can be ascending or descending curves, scale lines, arpeggios, and other simple figures that create their own natural pattern of anacrusis, focal point and release (Barra, 1983, p. 47). Some suggest that most melodic phrases rise and fall in pitch (Huron, 1996). The

[22] Ideas of underlying voice-leading were most explicitly developed by Schenker but have become integral to much music analysis.

melodic arch is not the only possible arch; others include those of texture, rhythm, dynamics, or tempo.

1.2.10.8 Length

Often phrases are mentioned as part of a list of terms for different lengths of musical units. For example, the term is 'used for short musical units of various lengths: a phrase is generally regarded as longer than a Motif but shorter than a Period' (Macy). Some discussions suggest that phrases can vary in length but many authors mention standard phrase lengths, usually four or eight bars (including, Koch, 1787, 1983 and Rothstein, 1989). Both Koch and Rothstein, describe, ways in which the phrase length can be changed but is still considered to be the original 'basic' length usually a 'four-bar' phrase. Temperley, assumes a constant number of notes (the actual number depends on the specific corpus, Temperley, 2001). For some theorists, such as Schenker, phrase length can be determined by the underlying principal tones.

In some of the experimental phrasing studies phrase length is determined by the experimenters (for example, Palmer and Krumhansl, 1987; 1987b) but not always (Deliège, 1998; Schaefer et al., 2004). The latter approach is followed here.

1.2.10.9 Metre, rhythm and the phrase

There has been confusion between grouping structure (a level of which is often referred to as the phrase) and metrical structure. A single hierarchical structure that captured both meter and grouping was proposed by Cooper and Meyer (1960). Groups of notes were classified in terms of rhythmic units borrowed from poetry, implying both segmentation and an accentual structure. These units related recursively with other units.

However, a consensus has emerged that metre and grouping are best regarded as independent structures (for example, Lerdahl and Jackendoff, 1987; Rothstein, 1989 and Temperley, 2001). Metre involves a framework of levels of beats, derived from rhythmic structures,[23] arranged in strong-weak relations. Metre refers to the rhythmic and (in 18th-19th century music) equal-length units that may be concluded from both temporal information and non-temporal information (e.g. harmony). Metre is start-weighted; the accent is on the first beat with a clear hierarchy within it (Lerdahl and Jackendoff, 1987 and Rothstein, 1989). Grouping (and phrasing) is a segmentation structure with no accentual implications. Phrasing is not necessarily regular or start-weighted. Its accents can come at the

[23] Rhythms are hierarchically grouped into bars (or tactus) forming the underlying metrical framework that can be developed or negated by the surface rhythms (Jones, 1992). The underlying or most pervasive structure is usually notated through the time signature. A direct relationship between rhythmic patterns and metrical perception has been shown and modelled computationally (for example, Steedman, 1977, Lerdahl and Jackendoff, 1987, Temperley 2001, and Spiro 2002).

start, middle, end or a combination of these, and can be affected by the metrical structure. This is not to say, however, that there is no interaction between them (Temperley, 2001, pp. 60-1). Phrase and metric boundaries often coincide, reinforcing the overall perceived grouping (Lerdahl and Jackendoff, 1987), and the notated time signature can affect phrase perception (Sundberg, 1988, p. 59). Metrical structure, whether perceived by the listener or through visual cues by experienced performers, seems to affect perceived phrasing.

1.2.10.10 Combinations and interdependence of different musical elements

Several music analysts stress the interdependence between musical elements contributing to phrasing (including Macpherson, 1912, pp. 9-13). Some go further, saying that all elements of music are involved (melody, rhythm, harmony, counterpoint, dynamics, articulation) and they are closely interwoven (such as, Neumann, 1993, p. 259).

The studies of all of the previous chapters contribute to the conclusion of the characteristics and rôle of musical features in phrase perception (chapter 12). The various features mentioned above are discussed in the context of the case-study pieces throughout the rest of the study in an attempt to identify which features and feature combinations are important for phrases, phrase-parts and phrase-types under different circumstances (chapters 13).

1.2.11 Rule base for phrase identification

Theorists have suggested rule bases or implemented computational algorithms that identify phrases. Rule bases can be constructed with different emphases – theoretical (Lerdahl and Jackendoff, 1987; Temperley, 2001) and statistical (Bod, 2002; Ferrand et al., 2002; 2003; Temperley, 2001), which may or may not be explicitly based on experimental results (chapter 8). In some of these studies, listener responses are used to confirm theoretically or statistically based hypotheses.

In this study, the process is reversed; having identified a relatively wide selection of musical elements (looking beyond purely monophonic music and the, by now traditional, instantiations of Gestalt principles) in advance, the participants' responses are used as part of the characterisation of the features, feature-combinations, phrase parts, phrase types and phrase combinations. These characteristics are then formalised as a rule base and algorithm (chapter 14). The rule base here is intended as the preliminary to one that can be implemented. Having a running program, at this stage, would be of limited value as there is currently no large, reliably annotated database of music of the genres investigated here for testing. The ideas that form the basis of this rule base are tested with a set of test pieces (chapter 15).

1.2.12 Summary

The above discussion shows the complexity of the subject matter and the large variety of approaches used to investigate it. Though the area of phrasing has been a topic of investigation for some time, there are many aspects that are treated separately if not in a contradictory manner in the different approaches (such as the differences in emphases on phrase lengths, and essential phrase features and who they are perceived by) and there are others that have not been treated at all or that may be more profitably treated from other perspectives (such as the ideas of phrase boundaries). A number of different methods have been developed but have never been employed together in order to enable direct comparison, and questions have remained as to how to analyse the data (leading to the need for the classification agreement analysis).

In this study, in order to investigate the areas discussed in the Introduction (section 1.1), a number of approaches are developed together with reference to the studies and ideas discussed above, forming a combined approach to the question of what contributes to the perception of musical phrases in examples of western classical music. Specific studies are returned to in chapter 8 in order to enable direct comparison between their approaches and the type of music and responses gathered in the current study.

In a wider context, insight in this subject area may relate to other (artistic) information. This investigation begins with an exploration of phrase perception in songs which contains many of the approaches and questions developed in the following chapters.

Research Laboratory at the Centre for Music and Science, Faculty of Music, University of Cambridge, 2003

Chapter 2

Introductory Study – Phrasing in Songs: The first downbeat

2.1 Introduction and aims

The discussion in the Introduction (chapter 1) included the notions of a single phrase structure, the relation between phrasing and breath, both physical and metaphorical, and the comparison of musical phrase structure with linguistic structure. An appropriate form through which to explore these notions seems to be the song. This introductory investigation of songs takes the above topics as starting points: for the general discussions of phrasing as presented by analysts; as marked by musicians on scores; as seen in performance contours (beat length and intensity); and its relation with musical features; and how these different aspects relate to each other. This study begins with relatively 'simple' examples that are also not too different from the rest of the examples of the repertoire studied in subsequent sections. The Lieder by Franz Schubert are the culmination of classical song composition. Morgengruß from *Die Schöne Müllerin* and Gefrorne Tränen from *Die Winterreise*, are investigated here.

2.2 Why songs?

Songs have advantages over other musical forms for an investigation of phrasing. In contrast to much instrumental music from the western classical repertoire, the phrase structure of songs is usually considered to be relatively simple; it is often constrained by the (usually pre-existing) text and physical breath and the vocal line is accompanied by a piano.

Although on this basis it seems likely that the phrase perception of a given song would be consistent, there may be room for variety. Text-structure is not always unambiguous (Vendler 1997) and musical structure does not always conform to it. Indeed, the converse is sometimes the case, with, for example, musical considerations necessitating word repetition. In some cases, text structure is described as being reflected in music while, in others the two can contradict each other (Barra 1983, p. 35). Furthermore, though in most cases the vocal line dominates and determines the structure of the accompaniment, the accompaniment may not always be structurally identical to the vocal line and more complex relationships may exist (see also Zbikowski 2002).

An additional reason for the investigation of phrase structure in songs is the assessment of the usefulness of the non-musical cue (the text) to annotate phrasing in songs. As there is currently no large database of western classical music with phrase annotations, if text structure is seen to be clearly representative of musical phrase structure, an automatic annotation tool that could be useful for phrase-related research could be developed. Even the annotated corpora that now exist, such as the vast Essen Folk-Song collection, which has been manually annotated, were apparently annotated without reference to the text and the texts are not even presented with the melodies, leading to some problems of interpretations of results (Bod 2001).

2.3 Approaches

This introductory investigation approaches the above aims and questions from several perspectives:

1. A 'theoretical' approach that investigates music analysts' discussions of these two Lieder by Schubert. Most of the discussions of these pieces take them as general examples of Lieder, including the aspect of phrasing.

2. A common way of annotating phrases on scores by performers is marking 'phrase-arcs' on the basis of internal hearing or playing. Here, musicians were asked to carry this out for the two Lieder. This limits the current investigation to those who can play and/or sing from score. Responses of listeners from a population with more diverse musical experience are investigated in chapter 3.

3. The musicians' definitions of the term 'musical phrase' and reasons for their phrase identification in the Lieder are discussed.

4. Tempo and intensity contours of different publicly available recorded performances are analysed.

5. Musical analysis of the pieces leads to identification of musical features. The distribution of the musical features is then compared with the phrases identified in 1-4, which are in turn compared with the text structure.

2.4 Schubert's Lieder

2.4.1 Roots and forms of Schubert's Lieder

Schubert's Lieder are based in the strophic folk song and the Classical symphonic tradition. For Schubert, the ideals of *Volksong* and strophic song were not aesthetic constraints but options among many expressive possibilities (Rosen 1997, p. 122). The songs are 'firmly grounded, in idiom and procedure, in the 'Viennese symphonic' period of music, say from 1770 to 1830' (Brown and Sams 1982, p. 86).

2.4.2 Relations between text and music; 'Schubert's "recreations in song"'[24]

In Schubert's Lieder there are two conflicting structural drives: the text structure dictates the musical one, and the song is a 'recomposition' which can include the modification of structure and is not completely reliant on the original text structure. In this study, the extent to which the musical structure reflects or "deviates from" the text structure is investigated.

Unlike in cases in which the text is the master of the music, the relationship between words and music in Schubert's Lieder is more equal. The process of recreation can lead to co-existing but different musical and poetic structures (Cone 1974, p. 19; Dürr 1982, p. 2). Rosen and Greene go further, saying that 'musical considerations take priority over meaning and prosody for Schubert' (Rosen 1997, p. 72) and that '[t]he metaphorical motifs that Schubert creates go a long way toward making the experience of his songs very different from that of the poem they set; they make it impossible to say that his music merely provides a way of declaiming a text. Nevertheless, the motifs are only one aspect of the musical experience. They are the material out of which the songs are built. One hears not only the motifs but also the forms or structures which they articulate' and Schubert fashions a new form for each song (Greene 1970, p. 183).[25]

---

[24] Dürr, 1982, p. 11.

[25] An example of Schubert's freedom in song structure and the relation with the text is his Wanderers Nachtlied (D. 768). The poem is free with varied line lengths, shifting accent patterns, an asymmetric rhyme scheme, and a sentence

2.5 The two Lieder: Morgengruß and Gefrorne Tränen

2.5.1 Morgengruß from *Die Schöne Müllerin*

Morgengruß as a text setting

Morgengruß is the first of a series of three strophic songs of courtship in *Die Schöne Müllerin* and is among the closest to the older strophic folk-song forms, being the simplest of the set. It sets four verses of the poem by Müller, all with the same musical material, so the music is not directly related to the specific meaning of each word or even the differing internal structure of the verses (the score and text are given in appendix 2).

'It is often noted in discussions about strophic song that variation in the performance of the successive stanzas can not only obviate monotony but also modify the musical sense so that it corresponds more closely to the progress of the poetry' (Cone, 1998, p. 116). Conversely, the text can affect the perception of the musical form and 'meaning' 'even when the musical surface displays no apparent irregularities. Often the shifting phraseology of the verbal text can induce fresh construals of the overall musical design' (Cone, 1998, p. 116). For example, in Morgengruss: '[t]he reiterated six-line pattern of the poem's four stanzas accommodates a shift in their grammatical and rhetorical structure. The principal division in the first stanza occurs between lines five and six, when the initial greeting and questions yield to the protagonists stated intention: "So muss ich wieder gehen."' The remaining three stanzas consist of two three-line sections. 'The second stanza contrasts the watchful lover with the awakening beloved; the third asks two questions; and the fourth turns from the beloved herself to the image of a singing lark and the love it symbolizes. The musical surface of the four stanzas, identical except for variants of rhythmic detail, reflects the strophic pattern of the poem' (Cone, 1998, p. 116).

Cone's figure (modified below) 'correlates the verbal and musical aspects of the strophe. A, B, and C represent the musical subdivisions' (1998, p. 116) and describes the structure of Morgengruss. He omits the prelude and coda, represents the piano interlude with the dash between A and B and the punctuation between B and C with the fermata. In the second line, 'the five bars of C are basically two, repeated wholly and then partially – as is the corresponding poetic verse.' The last line shows the fundamental harmonic structure: a move from I to V, an elaboration of the V, and a return to I (1998, p. 116-7).

---

structure that frequently overlaps the line divisions. The song has only 14 bars providing no scope for elaborate patterns of hypermeter, but, in keeping with the metric freedom of the poem, there is no regular 2 or even 1 bar punctuation (Salzer 1987, p. 17). In the songs analysed here the structure is not as free as in the Wanderers Nachtlied but is not always completely strict (Rothstein 1989).

	A	-	B 𝄐	C
Number of bars	6		4	5= 2+2+1
Text-line numbers	1-3		4-5	6 (with repetitions)
Harmonic structure	I-V		V	I
Bar numbers	4-10		11-15	15-21

'The two major points of division – the interlude after A and the fermata at the end of B – enable one to realize a musical distinction between two ways of articulating the poetry. The words of the first stanza suggest the musical reading (AB)C, in which section B is heard as extending the dominant close of section A. In contrast, the remaining stanzas imply A(BC), with B serving as a development before the reprise of the tonic and its thematic material. In performance, then, their fermatas would be understated in comparison with that of the first stanza. Perhaps further distinctions might be made by means of rhythmic inflections that contrast the fairly loose connection of the final line in the second stanza, 'ihr blauen Morgensterne', with the periodic sentence structure that grammatically binds the last line of the concluding stanza, 'die Liebe Leid und Sorgen', to what precedes it. There is an opportunity, moreover, to establish the finality of that stanza. Section C, based on a single line and its repetitions, consists melodically of three subphrases: a reiterated vocal descent from mediant to tonic, imitated by the piano and followed by a concluding elaboration of the mediant in the voice. By making the mediant the goal of the section (and thus of the entire strophe), the performers can obviate a sense of finality … In contrast they can achieve closure the last time around by emphasizing the descent to the tonic, in both voice and piano' (Cone, 1998, p. 117, see also Lewin, 1986, p. 349).

For Moore, because of the text, bars 14-17 should be sung without a breath for verses 3 and 4 but not in the others (1975, p. 28). The difference in text (one sentence as opposed to two) calls for differences in performance, which are possible because of the flexibility of the musical structure.

Musical characteristics of Morgengruß

Harmonically, this song does not move far – primarily to the dominant and back. There are only two clear perfect cadences, both when the piano is alone (bars 3 – 4 and 20 – 22). The rest of the cadences are onto the dominant with greater or lesser sense of stability: bars 9 – 10, echoed in bar 11 in the piano, and bars 14 – 15. The modified repeat of the first vocal line in bar 16 is the first time the melody is supported by a root position tonic. This becomes a pedal note, which continues until the end except for the dominant interruption of the cadence in bar 21.

The vocal line does not have a clear arrival on the tonic of the piece (as indicated by the key signature and the opening and final cadence) possibly reflecting the distance and weakness expressed in the text. Instead, the vocal line is based around the V (E) (first line), leads to the V/V (B) (second line), and stays there (third and fourth lines), leading to its dominant F#, and returns to the B in the vocal line and V in the accompaniment for the end of the fifth line. The final line is dominated by the E-C# sixth, arriving finally on the C# (third of A). However, the song is not finished here and the piano 'voice' takes over, finishing on the $\hat{5}$ (e) in the top and A in the bass. Even the final close is not strong (it would have been stronger if A was in the top part).[26] Even though the song does not move far harmonically, it does not follow 'typical' cadential progressions (V-I with $\hat{1}$ in the vocal line, see chapters 1 and 10) and strong closes in the way that many other pieces do. Other musical cues are, therefore, also necessary.

Lewin describes the changing, context-affected perceptions of the same progressions, notes and harmonies, emphasising that none are wrong or less important than any other. The context-free G^6 of bar 4 comes to be heard as a 'questioned' dominant of C major in bars 9-12. The questioning turns into serious doubt in bar 13 where G^6 is heard as IV^6 of D major (Lewin, 1986, p. 356).

The texture is primarily that of melody and accompaniment, except where there is a counter-melody in the piano part (bars 16^3-21). This counter-melody is a rhythmically simplified echo of the concurrent vocal line, and is shifted by a bar. There are also piano extensions (bars 10-11 and 21-23) and a piano introduction.

The rhyming scheme of the poem is aabccb, so the poem is naturally in two halves. This structure is followed in the musical setting with a long rest and slight change of theme between lines 3 and 4 of the poem (though keeping the rhythmic structure of the first line).

The overall structure of the text is A (aab - text lines 1,2,3) B (ccb - text lines 4,5,6). This is modified in the musical setting in two ways: the last line of each verse is immediately repeated, followed by a repetition of part of it. The structure of the song goes from the A B structure of the poem, to A B A': A (aab text lines 1,2,3), B (cc text lines 4,5), A' (bbb' text lines 6,6,6'), the three repeats of the last line being almost as long as the first three lines. All three renditions of the last line are modified versions of the setting of the first line of the poem. The text lines are all between 7 and 9 syllables in length. In the setting all, except the last repeat, are 6 (for the first) or 5.5 beats long. There are metric variations in the poem. The poem has varying numbers of syllables in the verses, requiring the addition or subtraction of notes and rhythms in order to maintain the overall musical repetition (Stein and Spillman 1996, p. 195).

---

[26] There may be several reasons for this lack of complete finality: 1) The importance of E in the piece almost gives it the status of another tonic, 2) This is a strophic song and the end is repeated four times, there cannot be a strong close every time. 3) The protagonist remains distant and 'unresolved' for the whole song. 4) This song is part of a cycle (though many of the rest do have strong closes). For a discussion of respective strengths of different cadences, see, for example, Caplin (2004) and chapter 1.

This is seen as a complex ABA' strophic setting by Stein and Spillman (1996, p. 194). There are various ambiguities in the B and A' sections enabling and encouraging repeated listening: the harmonically unstable B section has a sequence using a modal mixture including a half cadence that does not resolve, and a metric ambiguity in the A' section creates unresolved metric tension. The vocal line develops the opening vocal gesture of the A section and this two-bar phrase is imitated in the piano right hand. Within a two-bar phrase then, the vocal down-beat E in bar 16 is contradicted by a downbeat piano right hand E in bar 17 (Stein and Spillman, 1996, p. 194). Which is the correct downbeat of the two-bar phrase? This metric and phrase ambiguity continues to the end. The vocal line comprises four 2-bar phrases (16-17, 18-19 etc.). Both the vocal line and piano accompaniment end on a weak beat. As each verse concludes, the repeated hearing is welcomed to make sense of tonal ambiguities of the B section and metric confusion of the A' section (Stein and Spillman 1996, p. 195).

The vocal text lines begin on the following bar positions 4.666, 6.833, 8.833, 11.833, 13.833, 15.833, 17.833 (19.666/19.583) and end in beat two of bars 6, 8, 10, 13, 15, 17, 19 and 21.[27]

2.5.2 Gefrorne Tränen from *Die Winterreise*

In Gefrorne Tränen, Schubert 'makes the images aural as well as visual. We *hear* what the words invite us to *see* in our imaginations…we have a distinctive set of sounds, each recognizable as itself whenever it recurs, that function the same way a verbal metaphor does' (Greene 1970, p. 182). The piano's middle Cs, rhythmically isolated, gradually become an aural image for the teardrop freezing (Greene 1970, p. 182) (the score and text are given in appendix 2).

Gefrorne Tränen as a text setting

Gefrorne Tränen has only one verse and no repetition of material (through-composed). The original six-line poem has three pairs of rhyming couplets (aabbcc), each line being 13 syllables long. In the setting, the structure is modified in two main ways: there are repetitions of the second half of the second 'a' and the second 'c', and the whole of the cc couplet (both times with the repetition of the second half of the second 'c'). Unlike Morgengruß, there is only one set of words so the musical material can be more closely matched to every word in the poem. The piano part sometimes has a melody line, either when the voice has stopped or a very similar or identical one with it.

The first two phrases (setting the first couplet) are an example of (the common) antecedent – consequent phrases. Both phrases are related to each other; the first

[27] The clearest way to represent bars and beats on graphs is to show the beat as the proportion of the bar (for example, in a 4/4 bar, beat 3 would half way through the bar, 0.5). To avoid confusion, this is also used in the text.

phrase is open-ended and the second begins like the first but is, relative to it, closed and stable at its ending (Greene, 1970, p. 183-4). Each of the phrases has a partial rest or stop in the middle (on "fallen" and "entgangen"). This stop is even weaker than that on "ab". "Gerforne Tropfen fallen" is the antecedent to the consequent "Von meinen Wangen ab"; this line is antecedent, on a higher level, to the consequent "Ob es mir denn entgangen, Daß ich geweinet hab?" The two lines together could be antecedent, on a still higher level, to the following two lines heard together as consequent (Greene, 1970, p. 184).

The antecedent-consequent structure reveals the relationship between the two phrases. For example, subordinate clauses and the clause on which they depend are set as an antecedent – consequent pair, as are pairs of phrases in which the first introduces the direct quotation of the second. 'In most general terms, it can be said that the musical relation of antecedent to consequent has two different kinds of semantic effect on the text: First it divides the text into two parts (this is sometimes done grammatically as well). Second, it uses one part of the text to begin an idea, to set it into motion and the other part to complete the idea, to bring it to a relative close' (Greene 1970, p. 184). Having heard the music and text of the antecedent phrase, one expects a consequent phrase with a certain melodic contour, rhythmic shape, and harmonic cadence, and a text that will complete the idea that the music made incomplete at the end of the first phrase (Greene 1970, p. 184).

The vocal text lines begin on bars (repeats of parts of lines in brackets): 7.75, 12, (15.75), 20.75, 24.5, 29.75, 33.75, (37.75), 39.75, 43.75, (47.75) and end in the second beats of bars 11, 15, (17), 24, 28, and 33, in the third beats of 37 and 38, in bar 43.25, bar 47.5, (and in bar 49.25).

These discussions indicate some possible phrases, relationships between them and reasons for their identification which are returned to in the analysis of the results (section 2.7).

2.6 Methods

2.6.1 Musicians' phrasing study

Twenty-six musicians were asked to take scores, play them as much as they found necessary, and then to mark 'phrase arcs' clearly on the music (the starts and ends of the arcs were then taken as phrase starts (PS) and phrase ends (PE) in the analysis below). The task was intended to mirror, to some extent, the familiarity and tools (the score and an instrument) a performer has during their first encounter with a piece though the songs were presented without the words and without performance markings, only with key signature, time-signature, note length, note pitch and bar lines (as presented in appendix 2). The musicians were divided into two groups, each given the pieces in different formats, with different number of bars per line, and in one of three orders.[28] If they identified more than one phrasing possibility they were invited to give them all (though none did), and were told that they could provide any number of levels of phrasing. The musicians were not told anything about the pieces. The overwhelming majority of musicians did not recognise them though some did comment that they were songs.

The musicians were asked to return the music within two weeks along with two questionnaires, which they were asked to read and complete only after having completed the musical part of the study. One gathered information about their general musical experience and preferences while the other gathered information about the pieces (whether they knew them) and phrasing (what the definition of the term was to them and how they identified phrases). The musicians' musical experience ranged from practicing performers to music theorists and music psychologists (the questionnaires are given in appendix 3.2). The distribution of years of 'formal training' was:

Number of years of formal training	**% of musicians**
0 to 4	22.2
5 to 9	40.7
10 to 14	25.9
15 to 19	7.4
20 +	3.7

2.6.2 Performance contours

The tempo and intensity changes of publicly available recordings were analysed (for a list of recordings please see appendix 2.1). The tempo contours were obtained using the mustimer program (developed by Murray Allan, Winchester University). Mustimer takes taps on a normal computer keyboard as inputs and gives as output the rate of each tap in Beats Per Minute (BPM) and the time of tap since the first tap (in seconds), these can be used to calculate beat length which was the measure used for the following discussion. The data is collected by listening to the recording of the piece and tapping in time to the beat. When there are no note onsets on beats, an estimate was made as to the position of the 'missing' beats while listening. When all the data was collected, an average was taken of all the 'estimated beats' from the nearest true note onset until the last estimated beat before the next true note onset. In order to reduce the tapping error the tapping process for each piece was repeated six times. The three most similar tapping sequences were then averaged and the average sequence was then analysed. An ANOVA test for each set of three or more tapping sequences for each recording showed that there was no significant difference.

[28] This study was carried out once with these two pieces only and once with other pieces in addition (chapter 6) resulting in three "orders".

The intensity contours were obtained using an intensity recording program by Nick Collins with supercollider on a Mac which took as input the recorded sound files and gave as output intensity contours (in dB). Performance contours of each performance were studied and then compared to the listener responses to the same recordings. A more in-depth discussion of previous work and methods for analysis of performance features can be found in chapters 1 and 7.

2.7 Results

2.7.1 Verbal written responses

The musicians were asked to answer two questions about phrasing at the end of the experiment, one general and one specific: What, in your view, is the meaning of the term 'musical phrase'? and Please describe why you marked the phrase marks where you did. The responses can be grouped according to five categories: Section, Components, Boundary, What it isn't, and Language, which are presented in graphs 2.7.1.1 and 2.7.1.2, appendix 2. Most of the musicians gave a synonym for phrase included in the broad category of Section (including, unit, entity segment etc.). A small number mention the boundary between phrases (breath or pause). Most mention one or more musical features (including harmony, melody and rhythm). Some mention what it differs from (including motif and segment) and some make the linguistic comparison (such as describing the phrase as a means of punctuation). The proportion mentioning the linguistic connection is very small considering that these are songs (though the words were not provided).

2.7.2. Morgengruß

2.7.2.1 Musician's responses

Voices marked

The musicians were free to mark phrase arcs anywhere. Most marked the vocal part and the upper line of the accompaniment. As shown in graph 2.7.2.1, appendix 2, of those that do mark the accompaniment, there is also a very small minority that marks the right and the left hand parts separately. These, however, often have PSs and PEs in the same positions in both hands. This discussion therefore concentrates on the markings assigned to the vocal part and the upper line of the accompaniment, which do not coincide but fall in similar areas.

Even though this is a piece for solo voice and its accompaniment, the musicians often seem to see them as independent with respect to phrasing. However, there are some cases in which only one part was marked. It may be that when they do coincide, one part is completely dominated by the other (especially accompaniment by voice).

Effects of presentation (format and order) and musical experience

The results did not show significant effects of format or order of presentation or musical experience of the musicians.

ANOVA and T test results for difference of responses grouped according to musical experience and effects of presentation for Morgengruß, PS				
	Years formal training	**Years playing**	**Presentation order**	**Presentation format**
F	0.0665	0.1256	0.0614	
P	0.99	0.97	0.94	
t				0.87

Phrase markings on the vocal part

Phrase starts and ends: Graph 2.7.2.2, appendix 2 shows five clear PSs: 4.666 (96.2% musicians), 6.833 (88.5%), 11.833 (84.6%), 15.833 (88.5%), and 17. 833 (73.1%), and positions with much smaller responses around these. Additional positions have responses on and around them: 8.833 (61.5%), 13.833 (69.23%) and 19.583 (57.69%). Graph 2.7.2.2 also shows that, unlike the PS, the majority of PE identifications are spread over more than one beat, whether overlapping with or preceding the PS.

Phrases and sub-phrases: Graph 2.7.2.3, appendix 2 shows that out of the twenty-six musicians, six show both long and short arcs (phrases and sub-phrases). Two of these marked two layers of sub-phrases. All the shorter phrases fall within longer phrases (there is no overlapping). Moreover, almost all of the positions identified as sub-phrases by these musicians are also identified as 'phrases' by others (sometimes even those that chose sub-phrases elsewhere). The only exceptions are PS positions 12.5 and 14.5 both chosen by the same person. This musician wrote in his reasons for marking the phrase that this phrasing makes some standard patterns more 'stimulating'. Overall, however, only a small number of musicians chose to show sub-phrases. When musicians did mark sub-phrases they tended to do so only for a small number of positions in the piece. Moreover, most of these coincide with the main phrases as marked by the majority. For these reasons it is difficult to distinguish systematically between the two levels. Conversely, however, it seems that whilst the majority chose the common PS positions there is a subgroup of musicians who chose rarer PS and PE positions.

Phrase responses and text structure

Graph 2.7.2.4, appendix 2 shows that the majority of PS identifications coincide with text line starts though the proportion varies among the positions. Each of the text verses is structured slightly differently, with more or less clear divisions between each line (section 2.4.2). The majority of musicians choose PS at the positions that coincide with the starts of lines 1, 2, 4 and 6 (and the first repeat of

6). Fewer choose the start of lines 3 and 5. Musically, position 8.833 (the beginning of line 3), and position 19.333 (beginning of repetition of 2nd half of line 6) may be less popular because, being shorter and cadential in character, they are more like codettas (they are marked as sub-phrases by two musicians, graph 2.7.3.1.3). Position 13.833 may be less popular because it is a sequence of the previous phrase. These three weak positions are not preceded by rests, so it could be concluded that a rest is needed for a strong PS. However, there is no rest before 15.833 (beginning of line 6 for the first time) and, despite this, this position is chosen by a large majority of musicians.

In general there is a coincidence of the starts of text lines and phrases identified by musicians. In addition, the structure indicated by the rhyming scheme aab - cc - bbb (section 2.5.1) seems to be reflected in the responses. The first line of each section has the largest response which decreases for subsequent lines within the sections.

Comparison of analysts' descriptions and musicians' responses

The starts of Cones' sections A, B and C, (bars 4, 11 and 15) are chosen by the same musicians (section 2.5.1) indicating that these listeners have the same idea of the sections as Cone (and Lewin, who highlights only bar 11, section 2.5.1). Furthermore, the description of the mismatch between voice and accompaniment described by Stein and Spillman (1996, p. 194, section 2.5.1) is represented in the responses of the musicians who annotate both the voice and accompaniment. However, Moore's instruction for omitting the breath throughout bars 14-17 in verses 3 and 4 (Moore 1975, section 2.5.1 above) is not reflected in these responses, most identifying a PS on 15.833, contradicting the text structure (which was not known by the musicians) in certain verses.

2.7.2.2 Performance features

Tempo contours

Considering the 'simple' phrase structure of songs, all of the performances could be expected to be very similar. In general, graph 2.7.2.2.1, appendix 2 shows that the tempo contours of the different performances have peaks and troughs in very similar positions and the 'average' beat lengths are rather similar among performers (table 2.7.2.2.1).

Table 2.7.2.2.1: average beat length for the different performances

Performer	Average beat length (sec)
Ian Bostridge	1.087
Dietrich Fischer-Dieskau	1.063
Matthias Goerne	1.25
Giselle Vaillant	1.01

However, the degree of tempo variation within the piece differs between performances (ANOVA for the four performances shows a significant difference $F(3, 948) = 22.82$, $p < 0.0001$). This is mainly because of the difference between Goerne's performance and the rest (post-hoc linear contrast analysis, $F = 62$, $p < 0.0001$).

In general, the average tempi of the verses for each performance are similar and the differences are not significant, supporting Friberg and Battel's claim that repetitions of different sections are not varied (Friberg and Battel 2002), and contradicting Cone's claim of variation between verses (1998, p. 116, see section 2.5.1):

Table 2.7.2.2.2: comparison of the four verses of each performance ANOVA		
Performer	**F**	**P**
Ian Bostridge	0.7702	0.51
Dietrich Fischer-Dieskau	0.1843	0.91
Matthias Goerne	1.853	0.14
Giselle Vaillant	1.11	0.35

However, there are differences among verses, mainly in location and degree of ritardandi. The recording with the largest variation is that of Goerne. T tests indicate that the differences between the first and second, and the first and third verses are significant ($p < 0.05$). In the first verse, there is only one very large ritardando: 15.666 at the pause, though 19. 666, 21.333-21.666 and 22.666 have smaller changes. Of these, 15.666 and 19.666 are in the text and the rest are after the end of the vocal line, in the last two and a half bars of the end. In later verses these positions are strengthened and new positions are introduced. For example, in verse 2, positions 16.666, 17.666 and 18.666 are lengthened. The same positions are included in lengthenings in the last verse while in the third verse, slightly different positions are accentuated by lengthening: 17 and 17.666. For some positions the differences may relate to the text. However some, such as bars 21ff, occur after the text ends.

Intensity contours

Table 2.7.2.2.3, appendix 2 shows the positions or areas around which there in intensity decrease. In general, the positions are similar in the different performances.

Comparison of the intensity minima with the tempo contours shows that decreases in both often coincide. However, there are additional positions with only intensity decrease. In addition, positions of lower intensity in one verse are also lower intensity in others. The majority of these positions also coincide with the PEs identified by the musicians. In cases that they do not, these lower intensities usually occur only in one verse for one performer. For some positions (such as 11.666) there is lower intensity in all performances in every verse. However, there are some differences:

1. For some positions (6.666), some performers have a lower intensity consistently in every verse (in this case Bostridge and Goerne) while others do not have a lower intensity at all (Fischer-Diskau) or only once (Vaillant).
2. Other positions (8.666, 15.666) have lower intensity in the majority of the performances, but not in all verses. Sometimes, this may be related to the structure of the text. However, the verses for which this occurs are different and there can be several reasons for this variation in intensity among verses.
3. Others have lower intensity in one verse and performance (5.333, Vaillant).

These results indicate that there are a core group of positions with intensity decrease in a number of performances and majority of verses. There are other positions that are chosen less often, but still feature in some recordings, and there is a small group of positions that feature only once.

The results indicate that there seem to be a main group of PE positions that can be identified using tempo and intensity contours.

Comparison of performance contours with analysts' descriptions

The positions discussed in analyses and in writings on performance are among those accentuated by tempo or intensity change. This indicates that, in broad terms, these same positions are important from the perspective of phrasing. However, there is only one position in which there is coincidence between the 'detail' of analysts' words and the performance contours. Cone's principal division in the first stanza that occurs between the fifth and sixth text lines is indeed seen in the tempo contours of all the studies performances. However, this division continues in the performance of subsequent verses. Cone also explains that the subsequent verses consist of two, three-line sections. Neither contour type supports this; the division between lines 5 and 6 remains in all performances. Cone's further descriptions of the division of the relatively large-scale divisions of the same musical material in the different verses, are also not consistently supported by the performance contours. Many of the positions highlighted in performance are also identified by the musicians.

It seems, therefore, that the majority of musicians identify some phrase positions, and less identify others. Similarly, some analysts concentrate on the same (Cone) or different (Moore) positions. Comparison with the performance features shows that in some cases ritardandi and diminuendi coincide with analysts' discussions and musicians' responses. However, there are cases in which the changes in performance contours are less emphasised in the other domains.

2.7.3 Gefrorne Tränen

2.7.3.1 Musician's responses

Voices marked

As in Morgengruß, musicians were free to mark phrase arcs anywhere. They mainly annotated the vocal line and the right hand of the accompaniment but some also annotated the left hand separately. As graph 2.7.3.1.1, appendix 2 shows, unlike in the responses to Morgengruß, there are more annotations of, and more differences between, the annotations of the right and left hand, though there are many more positions of overlap of both PS and PE between voice and accompaniment. For the few musicians who do annotate all three parts separately, their markings (unlike those of Morgengruß) indicate that these parts are viewed as three independent parts. Even in a song, there seems to be clear three-part polyphony with each part sometimes having its own individual phrasing. The question of phrasing in polyphonic music is discussed further below (chapters 3 and 10-12).

Effects of presentation (format and order) and musical experience

The results did not show significant effects of format or order of presentation, or the musical experience of the musicians.

Table 2.7.3.1.2: ANOVA and t test results for difference of responses grouped according to musical experience and effects of presentation for Gefrorne Tränen PS

	Years formal training	Years playing	Presentation order	Presentation format
F	0.2689	0.2873	0.2537	
P	0.9	0.89	0.78	
t				0.81

Phrase markings on the vocal part

Phrase starts and ends: There are two PSs about which all musicians agree: 7.75 and 33.75, those on which most agree: 12, 15.75, 20.75, 29.75, 37.75, 39.75 and 43.75 and another group on which fewer agree: 22.75, 25.75, and 47.5. Somewhat fewer musicians identify: 9.75, 13.75, 21.75,24.75, 26.75,35.75 and 45.75 (graph 2.7.3.1.3, appendix 2).

As in the responses to Morgengruß, two main groups of musicians' interpretations are identified: those that include only the most commonly chosen positions and those that include also the less commonly chosen positions.

Phrases and sub-phrases: As graph 2.7.3.1.4, appendix 2 shows, eight musicians have both long and short arcs (phrases and sub-phrases). All the shorter phrases fall within longer ones (there is no overlapping). Moreover, almost all of the positions identified as sub-phrases by these musicians are also identified as 'phrases' by others (sometimes even those that chose sub-phrases elsewhere). The only exception is a sub-phrase PS at 24. As only a small number of musicians show sub-phrases and as most of these coincide with the main phrases as marked by the majority, it is difficult to distinguish systematically between the two levels. Conversely, it seems that while the majority chose the common PS positions, there is a subgroup of musicians who chose rarer PS and PE positions.

Phrase responses and text structure

As graph 2.7.3.1.5, appendix 2 shows, the most commonly chosen positions almost always coincide with text lines or sometimes the punctuation within them. However, there are some text lines that do not have a large response (especially 'Daß ihr erstaart' in bar 24.5) and there are PSs chosen by some musicians that do not coincide with the starts of text lines and do not follow internal punctuation marks (such as 25.75). These PSs, however, do have musical reasons that may be traced back to the text's content rather than its line structure or punctuation. For example, in bar 25.75, the text is about freezing, to ice, like the morning dew.

Like in Morgengruß, there some areas in which the responses are spread over two notes, such as 9.5 and 9.75. If the first text line of the verse is divided, the division has to occur on 9.75 and not on 9.5. However, the musical elements (arrival on the long note at the start of the bar with the repeated perfect cadence in the accompaniment) coincide with a PS for a minority of listeners on 9.5. Most of those that identify a PS in this area (a relatively small group), choose 9.75, coinciding with the text subdivision (which is unknown to the musicians).

Some musicians choose a PS on the upbeat to bar 23. The text couplet begins a bar earlier at the upbeat to bar 22 with the text 'Ei Tranen' (O tears) followed by a comma. The rest of the text line continues, and here, after the comma, a PS is identified by 60% of the musicians.

Sixty percent identify a PS at the upbeat to bar 27. This is in the middle of a text phrase. However, the music is a sequence, a third higher, of the music beginning at the upbeat to bar 22. In the earlier case, the music was identified as a PS, and coincided with the beginning of a text phrase. In bar 24.5, the text phrase starts half way through the bar. This is unusual in this song as most text phrases begin on the upbeat to the bar. A small number of musicians however, identify a PS here.

Comparison of analysts' descriptions and musicians' responses

The musicians' responses coincide in terms of location with Greene's description of the antecedent and consequent phrases. The shortest phrases (those on the 'lowest' sub-phrase level) coincide with the smallest responses (bars 9.75, discussed above, and 13.75) and the next level (bar 12), has more responses, though still has a lower response than the 'highest' level phrases mentioned by Greene (bars 7.75 and 15.75).

2.7.3.2 Performance contours

Tempo contours

Considering the 'simple' phrase structure of songs, the performances could be expected to be very similar. However, as shown by graph 2.7.3.2.1, appendix 2, each of the performances studied here has different characteristics and overall the difference between them is statistically significant ($F = 43.41$, $p < 0.0001$). Though the average beat length is rather similar (table 2.7.3.2.1), the degree, direction, and sometimes location of tempo variation change between performances.

Table 2.7.3.2.1: average beat length for the different performances	
Performer	**Average beat length (Sec)**
Pears	0.58
Fischer-Dieskau	0.61
Vaillant	0.63
Goerne	0.67

Matthias Goerne's recording has the greatest note lengthenings, usually lasting for one note. All of the positions coincide with musicians' responses, but there are response positions that are not reflected by the tempo contours. This may be because the phrasing is so clear from the musical characteristics at these positions that extra support from performance features is less crucial. Overall, the tempo contours for the different performances of the strophic Morgengruß are more similar than those of the through-composed Gefrorne Tränen.

Intensity contours

Like for the tempo contours, there seem to be more differences among the intensity contours for this piece than for the Morgengruß. Though the average intensity (and overall intensity range) for the different performances are quite similar, the degree, direction and sometimes location of intensity variation, change among performances.

Table 2.7.3.2.2, appendix 2 shows the positions or areas in the piece around which there are intensity changes. There are five positions for which all four performers have relatively low intensities. For the rest of the positions, at least two of the performers (often three) have low intensities in the same place.

All of the positions chosen by musicians are also areas of lower intensity in at least one performance. However, there are some positions (such as 8.75) for which there is relatively low intensity in the performances that are not identified by the musicians.

Comparison of performance contours with analysts' descriptions

Much like in Morgengruß, the analytical discussions mentioned above refer to some of the same positions highlighted in the performance contours. However, the more subtle relationship of relative strength of antecedent and consequent phrases (of bars 9 and 11, and 13 and 15) does not seem reflected in performance. In terms of intensity contours: two performers (Pears and Vaillant), have lower intensities at each of the four positions mentioned, but only in one pair (the bars 13 and 15 in Pears) is there the expected relationship of greater and lesser changes in intensity. In terms of tempo contours, Goerne's recording is the only one to have such a relation between these particular bars though other tempo changes in his recording are greater.

2.8 Discussion

2.8.1 Text structure and musical structure

The results of this introductory study indicate that text structure, or more specifically, the line structure, is often reflected in musical structures and at other times, the content of the text is reflected. Moreover, sometimes the same musical structure is set to several different text-structures.

The musicians were not given the text so their responses are to the music only. Despite this, there are some positions which all, or almost all, the musicians identified as PSs and these coincide with text line starts, indicating that the beginnings of the text lines are clearly reflected in the music as PSs. These positions are always preceded by, coincide with, or are followed by, specific musical features (section 2.8.5).

Sometimes, when the structure suggested by the content of the text is different from that indicated by its line or punctuation structure, both seem to be reflected in the music. Though musicians respond in both kinds of areas, in some cases, when there is a conflict between the two, the musical sections coinciding with the content seem to be more strongly reflected in the musicians' responses than the latter. This indicates that if poetic texts are to be used as a non-musical annotation, they have to be analysed both structurally, for example for new line-starts and internal punctuation, as well as semantically. The relationship between text and phrase structures identified here indicates that the use of text structure as an automatic annotation tool without further analysis could be problematic (see also section 2.8.2). However, there are now many text annotation models and implementations that may be applicable to this task.

2.8.2 Melody and accompaniment

Some musicians only marked the vocal part and others also marked the accompaniment. Of those that marked the accompaniment, there was also a minority that marked the right and the left hand parts separately. For Morgengruß, these often had PSs and PEs in the same positions. For Gefrorne Tränen, on the other hand most of the phrase arcs for the vocal line and accompaniment overlap. Overall, for the musicians, the vocal line is important while the importance of the phrasing of the accompaniment varies between and within pieces and musicians.

Schubert's Lieder are 're-compositions' of poetic texts, they are not a simple setting of a song with its accompaniment (section 2.4.2). These responses indicate that for some musicians, the vocal line dominates over the whole structure. For other musicians, the piano accompaniment has its own individual phrase structure. This indicates that an automatic text annotation model would not be enough for such songs. Moreover, this indicates that 'phrasing' of a piece of music can be dominated by one part, but it may be multi-layered, with more than one part having its own, individual phrase structure which sometimes coincides with the others. At positions where the phrase structure of the two parts coincides, the overall phrase structure is clearer.

2.8.3 Relation between theorists' discussions, musicians' responses and performance contours

The theorists discussed above sometimes concentrate on a relatively small number of phrases in these piece. Most of the positions mentioned by theorists were also identified by musicians and in the performance contours though not always in the same ways. For example, in the detailed discussions of Morgengruß, the analysts highlighted differences between the verses. However, the results of the study of the performance contours indicate that, although the same positions are prominent, the more detailed relationships between positions discussed by the analysts are not always clearly reflected in the performance contours.

2.8.4 Phrase start and end: Position and area

Many of the theorists in the fields discussed in the Introduction (such as Lerdahl and Jackendoff 1987; Palmer and Krumhansl 1987; Deliège 1998; Temperley 2001) seem to assume that a phrase begins on one note and ends on another. The results here indicate that although each musician may have marked just one position, these occur over a range. In this study of the Lieder, the range is relatively narrow. However, in other pieces the range can be longer (chapters 3 and 8).

2.8.5 Musical features

The above discussion has identified a number of musical features. In Morgengruß, five PSs were identified by the majority of musicians and were accentuated in at least one of the performers contours: bars 4.666, 6.833, 11.833, 15.833, and 17. 833. Each of these follows a rest, an inexact repeat (at least a repeat of the rhythmic pattern) and coincides with the start of a text line. As discussed above, positions 8.833, 13.833 and 19.333 are also chosen but by relatively few musicians. These 'weaker' positions are not preceded by rests, so it could be concluded that a rest is needed for a strong PS. However, as discussed above, there is no rest before 15.833 (beginning of line 6) and this position is chosen by a large majority of musicians.

A number of musical features coincide with the identified PSs and PEs. Some of these musical features are also present in areas without PS or PE responses. Graphs 2.8.5.1-4, appendix 2 show feature locations. Included in the feature graphs are also locations of tempo and intensity change. Below each feature graph, the musicians' PS and PE responses for voice and right hand are given. Table 2.8.5.1 highlights the musical features that occur with the PS chosen by musicians in this study.

Table 2.8.5.1: Summary of musical features at annotated PSs for Morgengruß

Bar	Musical features
4.666	beginning of vocal line, inexact repeat of opening, following long note
6.833	inexact repeat of opening, following rest
8.833	metrically parallel, same length as previous phrase, hint of plagal, following long note
11.833	rest, first of sequence section, inexact repeat (rhythmic), following rest
15.833	inexact repeat of first phrase (long), resolution of imperfect cadence to perfect, following long note
17.833	exact repeat (short term), following rest

In Gefrorne Tränen there are three types of musical features: some occur very often, such as pitch jumps, and sometimes coincide with PSs. Others, such as long notes and rests, also occur often and usually coincide with PS and PE responses (such as bars 20–22) even if the response is not of the majority of the musicians. Others occur less often and only do so when there is a majority response from the musicians. For example, all of those that have above 70% response have cadential progression, and/or voice-leading. These are not present in those positions that have below 70% response (see graphs 2.8.5.3-4). Table 2.8.5.2 highlights the musical features that occur with the PSs chosen by musicians in this study of Gefrorne Tränen.

Table 2.8.5.2: Summary of musical features at annotated PSs for Gefrorne Tränen

Bar	**Musical features**
*7.75	beginning of vocal line, following 'cadential progression'
*12	bar line, change in motive, following long note, rest, cadential preparation, voice-leading
*15.75	inexact repeat, pitch jumps in both accompaniment and voice, following long note, rest, cadential preparation and explicit voice-leading, bar position of first phrase
*20.75	change in motive, rest, long note, cadential progression resolves beat later, bar position of first phrase
21.75	inexact repeat, rest, bar position of first phrase
*22.75	rest, pitch jump, bar position of first phrase
*25.75	inexact repeat, following rest and long note, bar position of first phrase
26.75	inexact repeat, following pitch jump and rest, bar position of first phrase
*29.75	long note, rest, change in motive, bar position of first phrase
*33.75	pitch jump, long note, rest, change in motive, following voice-leading, bar position of first phrase
35.75	bar position of first phrase
*37.75	pitch jump, long note, following voice-leading, bar position of first phrase
*39.75	exact repeat, following increase in underlying rhythm, long note, bar position of first phrase
*43.75	exact repeat, following long note, rest, change in motive, voice-leading and cadential progression, bar position of first phrase
45.75	bar position of first phrase
*47.75	pitch jump, long note, change in motive, voice-leading, bar position of first phrase
50.75	following cadential progression, bar position of first phrase (accompaniment only)
* = Vocal PS with > 70 % response	

2.9 Summary

The results of this study of musicians' written phrase responses, performance contours, and analysts' discussions of these Lieder indicate the following:

Agreement between musicians' responses - There is high agreement between musicians' responses; there are some clear PS positions identified by most and a smaller number chosen by fewer musicians. Even with the notated music and trained musicians there can be different interpretations.

Voices marked - Musicians marked phrasing throughout the voice part. Some also marked the accompaniment. This may indicate that for some musicians phrasing is represented in or by the vocal part. For others, phrasing is in both, especially when they differ.

Performance contours - Overall, there are many similar positions of tempo and intensity change but there are also differences both in location and degree of tempo and intensity change. Contours of repetitions of different verses are, overall, very similar but different in their details (Cone 1998; Friberg and Battel 2002).

Musicians' responses, performance contours and analysts' discussions - For the main phrase boundaries identified by musicians, there was some change in degree of tempo, intensity or both. In comparison with analysts' discussions, again the same positions were highlighted. However, the details of degree did not coincide, for the most part, in these aspects of performance.

Words and music - The text structure (especially text line starts and punctuation but also the meaning of the words) may be used to locate some PSs. However, there are cases in which there is more than one possibility for the location of PSs and even in songs some phrases do not coincide with the voice part.

Phrases and musical features - Some musical features occur systematically at the positions that musicians chose as PSs and PEs. These musical features seem to be more varied, and their interrelationships more complex than those discussed as indicative of phrases in other studies (including, Deliège 1998; Ferrand, Nelson et al. 2002; Cambouropoulos 2003, discussed in chapter 8).

These themes are developed and investigated further in this study. The phrase and musical features are explored in a range of different pieces and musical settings – from the point of view of the listener hearing MIDI renditions and real performances, the analysis of performances and other score-based studies. In the following chapter, the aims, methods and general analysis of the listeners' responses to phrasing tasks are discussed with emphasis on the general characteristics of the responses, level of consistency, types of differences, and effects of the experimental procedure and listeners' musical training. This is important given previous debates and omissions, and also enables the subsequent concentration on the relation between musical and performance features, phrase parts and listeners responses in subsequent chapters.

Chapter 3

Listeners' 'Phrasing' Study - methods and results: A fore-phrase

3.1 Introduction
3.2 The music investigated
3.3 Listeners responses: methods of data gathering and analysing the responses

3.1 Introduction

The preliminary study introduced questions concerning the investigation of phrase definition and identification and a methodology involving score annotation (chapter 2). Here, the investigation is broadened: 1) the range of music is increased in terms of composers, genres and instrumentation, and 2) the music is presented aurally allowing both the investigation of listeners' responses and an increase in range of musical experience of the participants. In this chapter, the pieces, methodology, and a general description of the responses according to general aims is presented. More detailed investigations of the same responses follow in chapters 4, 8 and 10. Musicians' phrasing responses to scores of three of these pieces are presented in chapter 6.

This chapter presents the general aims, hypotheses, methods and results of the listening studies providing a general description and analysis of the listeners' responses to phrasing tasks. Emphasis is on the general characteristics of the responses, level of consistency, types of differences, and effects of the experimental procedure and listeners' musical training. Most of these have been problematic or omitted from previous studies (discussed further in chapter 8). This section also presents the data that forms part of the basis of the following chapters In chapter 10 the musical positions identified are analysed and compared to musical features and feature combinations. The music investigated is first described, bringing out some of the considerations and aims of the study.

3.2 The music investigated

3.2.1 Combinations of features in pieces

One of the aims of this study is to assess whether it is possible to relate musical features and feature combinations to listeners' phrase identifications and to assess whether features change in their importance depending on the feature combination. Therefore, pieces are chosen for their features and feature combinations and to exemplify different musical features that have been suggested as being important for phrasing. These include: harmonic motion (Rothstein, 1989), melodic contour, rhythm (Palmer and Krumhansl, 1987), metre, similarity (including, Cambouropoulos, 2001), and deliberate ambiguity (Lerdahl and Jackendoff, 1987).

3.2.2 Monophonic or polyphonic music

Previous experimental and computational studies on phrasing or segmentation have concentrated on monophonic music. In some cases (including, Deliège, 1998; some examples in Ferrand et al., 2003; and Schaefer et al., 2004) the examples are from monophonic pieces. In others (including, Palmer and Krumhansl, 1987, and the rest of the examples in Ferrand et al., 2003), a multi-voice piece is made into a monophonic one for the study. Though much information can be gleaned from such studies, this one aimed to look at western classical music in a way that is most similar to that in which this music is usually heard. Furthermore, studying pieces that are multi-voiced means that a larger range of musical features can be investigated, including explicit harmonic progressions and changes in texture, and brings with it the question of the function and relative importance of different voices. Therefore, both monophonic and polyphonic pieces were investigated.

3.2.3 Range of instruments, composers and styles

Most previous experimental studies of phrasing or segmentation have used a limited repertoire: folk songs (from the Essen Folk Song Collection), Bach's Fugue XX from the *Wohl-Temperiertes Klavier I*, Mozart's K. 331, Debussy's Syrinx, or Wagner's 'Die alte Weise' and often include a very few pieces. This study concentrates on western classical music and within this limited scope aims to explore music from a range of eras, styles and textures. Pieces written for different instruments and instrument combinations are used in order to investigate possible similarities and differences between them. A larger number of pieces than previously investigated in a single experimental study are used to begin to expand the corpus of pieces studied with respect to phrasing.

It is not possible to test a large number, and therefore range, of pieces with each listener. One approach is to concentrate on just one genre (as in chapter 2). The current one is to use a wider range of music and to treat each piece as a glimpse of a larger corpus of possibly similar pieces. Investigating pieces by composers of the 'canon' also implies that the pieces have been regarded as 'successful'.

3.2.4 The Pieces

The pieces are from the Baroque, Classical, Early and Late Romantic eras and were chosen primarily for their feature combinations and the nature of their phrasing (from the straightforward to the more complex). Here a brief overview of some of the features and phrase elements that prompted the choice of the pieces is given. These will be discussed in more detail in chapter 10.

The pieces were chosen from several genres including solo music and arias for voice and orchestra. Solo instrumental music is not restricted by a text. For non-wind instruments, the phrase lengths are not limited by 'physical' breath, possibly allowing for longer, more complex phrases, and there is no need for consideration of other instruments. The aria allow for the investigation of 'polyphonic' music with differing relationships between the vocal and orchestral parts. The excerpts are presented in appendix 3.1.

J.S. Bach, *Matthäus Passion, Soprano Solo Aria, 'Wiewohl mein Herz in Tränen schwimmt'* [henceforth, **Bach Passion**]

In this excerpt, the harmonic progressions are unusual and often unexpected; the soprano part contains many characteristics that are not typical of a 'stereotypical sung melody'; the piece is written as a solo and accompaniment. However, the phrasing of the vocal line and the accompaniment do not seem to coincide enabling the discussion of the relationship between different musical lines.

J.S. Bach, '*Cello Suite no. 5, Prelude* bb. 1-7.5 [henceforth, **Bach Suite**]

In this excerpt, there are harmonic and rhythmic 'resolutions' that are prolonged, the metrical structure is somewhat ambiguous; the phase changes so that the third beat of the bar sounds like the first, and the piece is mainly monophonic.

W.A. Mozart, *Piano Sonata, K. 310: Andante cantabile con espressione. bb. 1-8* [henceforth, **Mozart Sonata**]

Here, the phrase structure, though not completely unambiguous is, in some ways, straightforward. There are only four potential phrase boundary areas. Several aspects are less clear including the exact position of two phrase boundaries. For example, the main phrase boundary in bar 4 in theory can be at two places and that in bar 6 is somewhat ambiguous. It may be seen as an elision, or the previous phrase may be seen as ending on one semiquaver and the next may start on the following (weakly positioned) one.

W.A. Mozart, *Le Nozze di Figaro, Act IV, No. 26 Basilio Aria. bb. 106.5 – 118* [henceforth, **Mozart Aria**]

In this excerpt, the phrase structure is relatively clear. The main question is the phrase "level" identification; two sets of four bars have a constant harmony and pitch class. The relationship between the vocal line and the accompaniment is more straightforward than in the Bach Passion.

J. Brahms, *Intermezzo, Op. 119, No. 1. bb. 1- 16* [henceforth, **Brahms**]

In this excerpt, certain phrase parts, which may usually be regarded as essential, are problematic. For example, the only certainty that the phrase end has been reached is provided by the start of the next. At the same time, phrase starts may be expected several times before one arrives. This is because of the presence of different features (including melodic descent, change in texture, repetition) and enables the exploration of which musical features have a stronger effect on listeners' phrase perception. The harmonic structure is difficult, partly because the tonic remains unclear throughout.

R. Wagner, Cor Anglais Solo, 'Die alte Weise' from Act III of *Tristan und Isolde* [henceforth, **Wagner**]

This excerpt is completely monophonic. The harmonic structure is very difficult and the voice-leading forms implications but ventures further than expected. The piece was used by Deliège (1998) so the two studies' results can be compared.

3.2.5 Piece-length

The whole of the relatively short Bach Passion movement and self-contained section of the Wagner were played. Because of time constraints, relatively short excerpts of other pieces (though longer than in many other studies) were presented. For the Bach Suite, Mozart Sonata and Brahms the opening of the pieces, where precedents for the rest of the piece can be set, were played. The Mozart Aria excerpt is from the middle of a movement. The excerpt lengths may affect the phrases identified (chapter 10.8).

3.3 Listeners responses: methods of data gathering and analysing the responses

Here the general aims of the listeners' study are discussed (A.) A brief method follows (more detailed methods follow the hypotheses in appendix 3.4). This is followed fir a presentation and discussion of the results (R and S sections respectively).

3.3.1 General Aims

The general aims of this chapter can be framed in the form of the following questions:

A.1. What is the nature of the written, verbal and musical responses?
A.2. Which sections of music are identified by listeners as phrases and what are their characteristics?
A.3. Are there differences between the responses related to musical experience?
A.4. Does prior familiarity with the piece have an effect on responses?
A.5. Are there differences between consecutive listenings by the same listener. If yes, what are they, and do they indicate learning?
A.6. Are there variations between listeners' responses affected by the order of pieces or tasks carried out and if so what are the implications?
A.7. Are there differences between responses to MIDI and performance renditions of the same pieces?
The following two aims will be discussed in chapter 10
A.8. Do the identified phrases coincide with those predicted by music and music-psychological theories?
A.9. Do these 'phrases' coincide with the presence of musical features?

Aims 1 and 2 address the core of the subject of this study regarding the perception of phrases by listeners. Aims 3 – 7 address details of the responses and their diversity or consistency in the context of personal and experimental factors. Aims 8 and 9 address other core aspects of this study and will be discussed after the music and music-psychological theories and the musical features have been identified and analysed**.**

A. 1. What is the nature of the written, verbal and musical responses?

This includes a number of aims: To determine whether a) the task was clear and the term 'phrase' was self-evident to the listeners; b) there was a clear theoretical meaning of the term for the listeners and if there was, what it was and whether it was related to musical experience, c) there was a systematic response to phrase part identification (phrase starts, phrase ends, and beginning of the expectation of the phrase end).

A. 2. Which sections of music are identified by listeners as phrases and what are their characteristics?

Listeners' phrase identification is the critical step of this study and the aim was to characterise the phrases identified, test whether there was a variation in their nature and to investigate the possibility of identifying musical reasons for those decisions.

A. 3. Are there differences between the responses related to musical experience?

Some previous studies analysed responses by listeners with different musical experience (chapter 8). Some theories assume that there are differences depending on level of experience (Lerdahl and Jackendoff, 1987), others concluded that there is little effect of musical training (Deliège, 1998) and others conclude that there is (Schaefer et al. 2004). Here, the extent and types of differences between responses by listeners with different musical experience is assessed.

A. 4. Does prior familiarity with the piece affect responses?

Getting to know a piece involves learning its themes and structure and this may be in part helped by the memory of the phrases. The effect of familiarity with the piece is explored.

A. 5. Are there differences between consecutive listenings by the same listener. If yes, what are they, and do they indicate learning?

In some previous studies listeners were given training sessions before the responses were recorded and in others the first listening(s) were treated as a familiarisation runs in which the subjects were not requested to give a response, thus allowing for 'learning' but not recording the process (if indeed it is occurring) (e.g. Palmer and Krumhansl, 1987, Deliège, 1998, and Schaefer et al., 2004). Here, responses to the first listening were recorded in order to obtain the immediate, online reaction. These listenings, were treated as reflecting most closely 'normal' listening to a new piece, in which case listeners do not know what is coming next, so reactions are based on incoming information and possibly predictions. This was followed by two further listenings in order to asses whether multiple hearings (and therefore further familiarity with the piece with the same task in mind) affect phrase identification.

A. 6. Are variations between listeners' responses affected by the order of pieces or tasks carried out and if so what are the implications?

The effects of piece and task order were investigated.

A. 7. Are there differences between responses to MIDI and performance renditions of the same pieces?

Performance features may be necessary for listeners to identify phrases.

A. 8. Do the identified phrases coincide with those predicted by other theories or discussed by other theorists?

Another aim was to compare the sections identified by listeners with those identified by theorists and then to compare these to the use of the term in the musicological, music-psychological and computational literature to identify similarities or differences on a theoretical level (chapter 10).

A. 9. Do these segments coincide with the presence of features?

One of the main aims was to investigate whether the phrases identified by listeners coincide with musical features and to then describe the functions of different features and feature combinations (chapters 10 and 11). Hypotheses relating to each of these aims are given in appendix 3.4.

3.3.2 Method

3.3.2.1 Tasks

There were two tasks:
1) Phrasing: the listeners were asked to press a key at the beginning of a phrase (PS), hold their finger down for the duration of the phrase and lift their finger at the end of the phrase (PE).
2) Beginning of the expectation of the end (EOP): the listeners were asked to press a key when they began to expect the end of the phrase (EOP) and then lift it again and press again if the phrase did not end where expected (see appendix 3.2 for procedure).

The phrasing and expectation tasks were given at least three times to each listener and responses were recorded from the first listening. Listeners could repeat the tasks as many additional times as they liked (two was the maximum). In this analysis, three listenings were taken: first, middle and final.

3.3.2.2 Stimuli

The MIDI stimuli were prepared manually in MIDI format on Digital Performer 3 on a Macintosh computer. The MIDI rendition allows the 'nominal' or deadpan performance i.e. a direct translation of the score into physical variables, where all notes of the pitch and inter-onset-intervals of the score (Friberg and Battel, 2002, p. 201) and the intensity is equal throughout. The MIDI renditions used 'piano sounds' for all the pieces originally written for piano and for the opening of the Bach Prelude. For the Wagner a 'clarinet' sound was used and for the two vocal pieces, piano was used for the accompaniment and 'clarinet' (without words) for the voice.

Publicly available recordings of performances that were as different from each other as was available were used in the second study (a list of recordings is in appendix 3.3). The stimuli were prepared on Digital Performer 3 on a Macintosh by importing performances from CD recordings. There were two sessions (I and II), and the listeners were split into two groups. The first group heard one set of recordings in session I, and the other in session II and the second group heard the groups of recordings in reverse order. For more details of the methodology see appendix 3.4.

3.3.2.3 Listeners

Listeners were mostly students from the University of Cambridge. For the MIDI study of the total of 35 listeners, 20 were music students or held a music degree (degree-level musicians - DL), 5 had played an instrument regularly in the last 10 years (musicians - M) and 10 listened to music of some kind but did not have any formal training (non-musicians - N). The M and N are, for some of the analyses, treated together (non-DL). For the study of responses to performances (six months after the MIDI study) there were 48 listeners (26 DLs, 11 Ms and 11 Ns), 15 of whom had participated in the MIDI study.

3.3.3 Results

R.1. What is the nature of the written, verbal and musical responses?

R.1.1 Is there a response to the definition question and the response?

Almost all of the listeners 'knew' what the term meant (to them) from the start of the session. For those that did not, after the analogy with language and the example of the nursery rhyme had been given, they said that they understood. Some pieces were more difficult than others but all the listeners provided responses for all tasks and questions.

The written (R.1.2.a-c) and verbal (R.1.2.d) definitions and descriptions from all three listening sessions are discussed and compared with the terms used in the literature (R.1.2.e). This is followed by a general discussion of phrase identification responses.

R.1.2.a Written definitions following listening to MIDI renditions

In the answers to the question: 'What, in your view, is the meaning of the term 'musical phrase'?' key terms were identified (of which there were usually more than one in each response). Not surprisingly, considering the freedom of the question and the small sample, each term was repeated relatively few times. The terms were analysed and the following categories emerged.

Categories	Terms
Section, part	'Clause', 'Section', 'Chunk', 'Entity' ('open'/'closed'), 'Segment' (DL), 'Bit', 'Short part', 'Building block' (Non-DL), 'Unit' (Both).
Linguistic references	'Sentence', 'Clause'
Components, features	'Melody', 'Harmony', 'Rhythm', 'Motive', 'Texture', 'Related sounds and rhythms'
Boundary	'Breathing point' (DL), 'Pause' (non-DL), 'Break' (both).
Ending (and Beginning)	'Sense of arrival', 'resolution/closure', 'Closure', 'From logical beginning until cadence', 'Comes to rest somewhere logical' (DL), 'Musical thought developed and brought to resting point', 'verse ends' (non-DL).

The categorization is not sharp or mutually exclusive. However, it gives an overall idea of the important explicit elements in the perception of phrases, indicating possible musical components used in phrase identification. The proportion of terms within each category was calculated as the frequency of the use of terms within that category out of the total number of occurrences of all terms (graph R.1.2.a, appendix 3.5)

The majority of the categories include terms given by at least one DL and non-DL, suggesting that there was no general division according to training. However, a variety of individual terms are used and even if more than one listener uses the same term, it is usually used only either by DLs or non-DLs with little overlap. In the table, the main categories are listed according to their frequency.

The list of categories and terms indicates the range of aspects associated with the term phrase. Listeners with different musical experience choose different synonyms. Only 'unit' is used by both DL and non-DL and is the most common term in the subgroup (4 listeners). Both DL and non-DL use terms that refer to sentences, clauses or phrases in language or the punctuation between them within a broader definition. DLs dominate the components and features group. Even for the majority of DL, general definitions do not include individual features and only harmony is elaborated (harmonic goals and cadences). Given the task of providing a (short) definition, clear functions of individual features are not part of the overall definition. The boundary category includes contributions from DLs and non-DLs. Breath is also used by DLs describing the 'whole', both as a metaphor and as a practical limitation. There is a concentration on phrase ends over starts, with only one mention of beginnings: 'from a logical beginning until a cadence, (imperfect/perfect)' (DL). This bias could be because of the tasks carried out, especially the EOP. However, listeners were also asked to identify phrase starts. 'Boundary' and 'Ending' terms are used by different listeners. Only one (non-DL) uses both. DLs and non-DLs are represented in both categories, though for the 'endings' there is a greater proportion of DLs and for the 'boundaries' there is a larger proportion of non-DL. All but one mention either boundaries or endings.

The rest of the terms are presented in the table and discussion in R. 1.2.a, appendix 3.5.

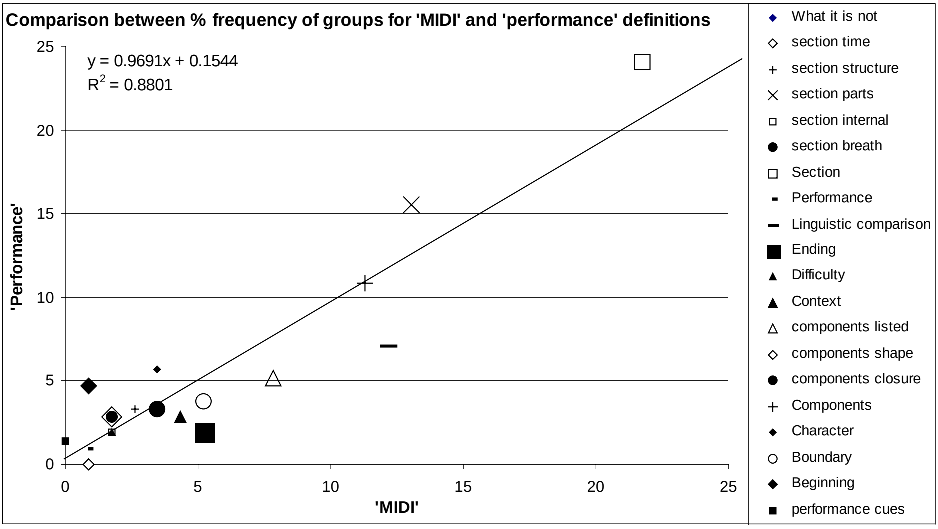
R.1.2.b Written definitions following listening to Performances

Terms for MIDI and Performances

Fifteen listeners provided written definitions after both sessions, which are included here and are also discussed separately in the following section (see R.1.2.c). On average, there are 3.7 terms per person for MIDI and 4.7 for 'performance'. However, the variation in length of responses is very large.

The same analysis was carried out for the responses to the same question at the end of the performance session and showed similarity with the above responses. It is possible to include these terms in the same categories. The frequency of the occurrence of the terms belonging to the respective categories is very similar (graph R.1.2.b.1).

Graph R.1.2.b.1



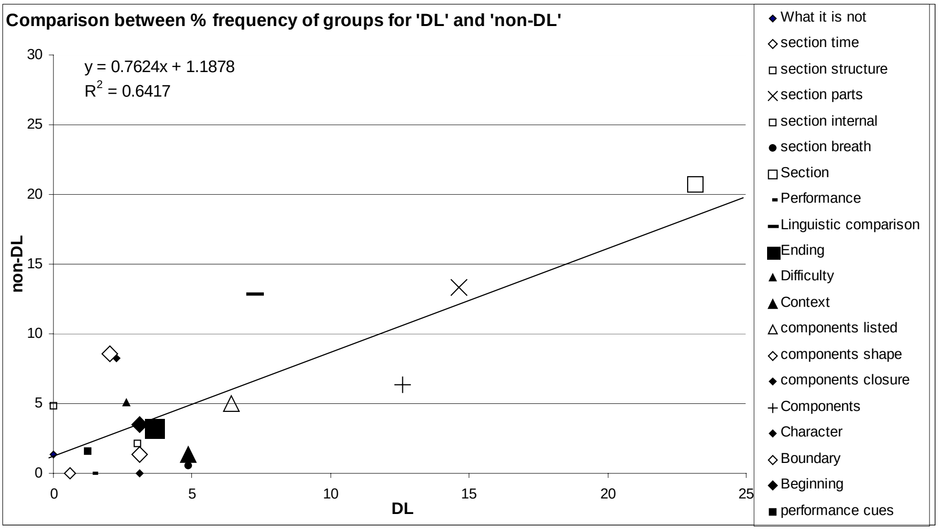
Graph R.1.2.b.1 shows the same distinction between the terms used frequently (>6%) and infrequently (<6%) for both MIDI and performance sessions. The high frequency categories are: *section, part, linguistic comparison, components* and *components listed*. *Section* and *components* include several subcategories including *part* and *components listed*. This distribution reflects the tendency of proceeding from comprehensive to detailed terminology and different listeners proceeded to different degrees. The use of specific terms of a 'detailed' nature reflects the importance attributed by the listeners to these terms as essential for the definition of the term.

Two categories occur in the 'MIDI' but not 'performance' responses: *what it is not* and *section, time*. The single occurrence of 'musical time' is interesting considering its importance in the music-psychology literature. Conversely, the only terms mentioned in the 'performance' and not in the MIDI responses are specific *performance cues* (by two non-DLs). It is surprising that this category of terms is not more frequent especially following the performances.

Terms used by DL and non-DL

Graph R.1.2.b.2, comparing the DL and non-DL responses, shows that the categories of terms used by the groups overlap.

Graph R.1.2.b.2



The most popular terms overall are equally represented in the two populations, other categories of terms are relatively frequent but differ between the two groups, others are relatively infrequent (<6%) in both populations, for some the frequency is similar and for others different between populations.

R.1.2.c Consistency between MIDI and performance sessions

Fifteen listeners gave definitions both in the 'MIDI' and 'performance' sessions. These were compared according to four criteria (see R.1.2.c in appendix 3.5 for examples and their discussion). In the definitions provided after the performance session only four included terms such as 'played', 'sung' and 'performance'. In two of these, they were in the context of breath and in one as a part of a more general statement. Only in one is there any mention of a 'performance feature'. Even this statement is not completely unambiguous; this listener may have been referring to longer notes or to overall tempo reduction (see R.1.2.c).

The definitions given after listening to MIDI and Performance renditions broadly include the same terms and ideas. However, they differ in their extent and details. The definitions did not seem to be directly influenced by the type of rendition, and showed no large differences between DL and non-DL.

Summary of written responses

That everyone was able to give a definition for the term 'phrase', indicates

- That there was a general understanding of the term,
- That the term was meaningful to all listeners, and
- What it meant to each one.

Listeners provided a large variety of terms, which are used in a number of different disciplines, including music theory and analysis, linguistics, psychology and have computational, physical-biological connotations. There does not seem a direct link between the types of terms and the listener's musical experience, indicating that the meaning assigned to 'phrase' links with a broad, common and inclusive vocabulary.

The individual terms differ slightly between DL and non-DL indicating that musical training affects terms used. However, there is great commonality between the categories the terms belong to for both groups. The general ideas associated with the term 'phrase' seem to be similar regardless of training.

The comparison of responses of the MIDI and 'performance' sessions also shows a remarkable lack of difference between the responses. Moreover, only two listeners include 'performance features' in their definition after the performance session and none does so after the MIDI rendition; they did not consciously use them in the performances and did not notice their lack in the MIDI. Indeed, the performance features did not form a part of the (written) definition of the 'phrase' for the overwhelming majority of listeners.

R 1.2.d Terms used in verbal description of what influenced the decision of listeners on phrase boundaries and in their comments on difficulties encountered in this task.

At the end of each task, listeners were asked to explain any difficulties, and what made them put the phrase marks where they did. The complete list of terms is given in table R 1.2.d, appendix 3.5 and the groups of terms are discussed in section R.1.2.d, appendix 3.5). The responses in both the MIDI and performance sessions include a variety of terms from the general (such as 'shape', and comparison with linguistic ideas) to the specific (such as musical features and phrase parts), on which there is much more concentration. These musical features also range from the general (e.g. harmonic) to the more specific (e.g. perfect cadence).

Though there may be a basic 'list' of ideas used by listeners, indicated by the frequent comment on 'missing' features (such as lack of clear harmonic progressions in the Wagner), the terms used vary between and within the pieces, suggesting that listeners are flexible when listening. Listeners, especially when responding to the Mozart Aria, Brahms and Wagner, comment on changing usefulness of different characteristics within the piece. In general, in responses to both MIDI and the performances, the harder the listeners found the piece, the more they talked about it afterwards.

Comparison between DL and non-DL and MIDI and performance

In the MIDI responses there seems to be a broad distinction between DLs and non-DL: DLs tend to mention harmonic features and structure (regular or irregular). Non-DLs concentrate on rests, pauses, pitch, and the unexpected nature of some phrases.

There seems to be more variety of categories between pieces and performances in the performance responses. In general, harmony is still often mentioned among the DLs, and in several of the pieces there is comment on the difficulty of the overlapping phrases or continuous feel of the piece. Beyond these, the popular categories and the group mentioning them change depending on the piece. In some pieces, some categories are mentioned for both renditions by DLs and non-DLs (performance cues in the Bach Passion, or rests and pauses for the Mozart Sonata performances). Other categories are mentioned only by one group for both renditions (change and pitch for non-DL in both renditions of the Mozart Sonata) and others only by one group for one rendition (regular structure by DLs for the Uchida, beginning by non-DL for the Gould). There are also some categories mentioned for the different renditions by different groups (performance features by DLs listening to Uchida and non-DL listening to Lipatti).

DLs and non-DL comment equally on difficulties encountered when responding to the performances though sometimes these are described differently. Relatively few listeners comment on the lack of performance features in the MIDI (only one mention in the Bach Suite and Wagner). Some listeners even mention performance features (getting quieter or slowing down) for the MIDI. This is reminiscent of studies in which such performance features are perceived or used even if they are absent such as isochronous tones heard as being different in length (Abecasis et al., 2004), or performers asked to perform music "deadpan" still add temporal and intensity variations (Drake and Palmer, 1993).

For the performances, there is a rather large reference by listeners to performance features in all pieces (except the Mozart Aria). However, there is also a surprisingly large proportion of listeners who find the performance features contradict other 'musical' features or make it more difficult to identify phrases (Bach Suite, Bach Passion, Wagner, Brahms). Moreover, when such contradictions are mentioned, listeners tend to prefer the other features to the performance ones. This indicates that these listeners have an idea of the phrase structure that is 'separate' from the one presented by the performer, suggesting that when listening to performances at least some listeners are not purely reliant on the performance features but have their own interpretation and expectations based on other musical features. Those mentioning the performance features in this way included listeners who had not heard the MIDI renditions before and did not know the pieces. Several listeners, when making these comments, then made value judgements about the performances in general (usually negative). These features that affect phrase perception may also possibly be influencing the extent to which the listener likes a particular performance.

R. 1.2.e Comparison with terms used in examples from the literature

The meanings of the term 'phrase' given in the literature discussed in more detail in chapter 8 are presented and then compared with the different 'descriptive' and 'operational' definitions given by the listeners in this study.[29]

There are some differences between the context of the definitions described so far and those collated from the previous studies: in several cases the 'phrase' is mentioned as part of a study that is not specifically about 'phrases' but about 'grouping' or 'segmentation'; several of the studies consider only monophonic music and some consider specifically 'melodic phrase structure'. There does not seem to be a clear distinction between operational 'how is the phrase recognised' and descriptive 'what is it' definitions, so in this discussion all are considered together. It is noteworthy that no 'functional' definition of a phrase types 'what different phrases do' occurs in the literature reviewed. Although there are cases in which some aspects may be inferred from the musical examples and those will be discussed in chapter 10, here only explicit definitions or descriptions are included.

The aspect that recurs most often in the literature studied is that the phrase is a unit, segment or major articulation. These units are often described as part of a hierarchy of these or other types of units such as the shorter motive or the longer periods. The definitions often seem rather vague, for example with reference to the 'position' of this unit within the hierarchy (e.g. Temperley, 2001). However, when there are explicit descriptions, there can be (unacknowledged) contradictions between different studies. For example, for Lerdahl and Jackendoff the 'phrase' is relatively high in the hierarchy, while for Cambouropoulos it is relatively low (chapter 10).

The main aspect of 'phrasing' concentrated on is phrase boundaries and their characteristics (such as Cambouropoulos, 2001, 2003, explicitly, or Lerdahl and Jackendoff, 1987, implicitly). Temperley differs from others by distinguishing

[29] The term is more widely used than is reflected in the comparison carried out here.

explicitly between phrase ends and starts in his musical examples (rests are not included in the phrase), however, his program identifies only phrase starts.

Several theorists (such as Lerdahl and Jackendoff, 1987, and Deliège, 1998) and those that relate their work to these, refer to grouping and the gestalt principles (chapter 1, section 1.2.8) though the exact relation between types of grouping and phrase is not clear. Some, such as Lerdahl and Jackendoff (1987), make reference to language (but there is no explicit comparison between phrases and sentences). The aspects included in the different phrase 'definitions' are listed in table R. 1.2.e appendix 3.5, along with the categories identified from the responses above and 'new' categories mentioned only in the literature.

Several aspects of the listeners' definitions overlap with those discussed in the literature reviewed (table R. 1.2.e, appendix 3.5). For example, the prevalence of synonyms of unit and some of the categories that emerged in this study relate to common gestalt principles (change, long notes). However, some aspects occur only in one group and there are differences in the 'difficulties' (table R. 1.2.e, appendix 3.5).

R.1.2 Summary of written and verbal responses

The written definitions and verbal discussions indicate that:

- The definitions and associations vary, but that they fall into common categories – synonyms for the terms phrase concerning units, 'music-theoretic' characteristics (e.g. harmony), characteristics that can be related to Gestalt ideas (change, repetition etc.), and physical aspects (breath).
- There is some overlap between DL and non-DL responses, though sometimes the terms are more music-theoretic in the case of the DL.
- There is an overlap between the terms included in the listeners' responses in this study and those used in the literature. However, not surprisingly, each study concentrates only on one or two of the aspects listed above. Moreover, in the literature the concentration seems to be on the boundary areas, while in the responses, especially reflecting the difficulties encountered, the structure of the phrase (especially in terms of the build up of expectations) seems to be important. In this study the instructions included a mention of expectation.
- Overall, many (though not all) of the terms used both in the literature and responses remain broad or even unclear. This study investigates the musical cues and aims to clarify the definition of the term phrase and the rôle of the musical cues in its identification.

R.1.3 The nature of the responses to MIDI and performance

The results of the MIDI and performance phrasing studies are presented in the form of response profiles showing the % response for every position for all of the responses for the phrase starts (PS), phrase ends (PE) and the beginning of the expectation of the end of the phrase (EOP) tasks (first four graphs in each appendix 3.6.1-6). The time lines are presented according to beat position in the bar (discussed in appendix 3.6.A, time unit representation).

The graphs show that rather than specific positions, there are areas of responses for the PS/PE and EOP with broader or narrower peaks. In some cases, the peaks of EOP are lower but broader than the PS/PE. In most cases, there are areas of high response in all three measures and other areas of low response indicating that there are areas preferred by a large proportion of listeners. This indicates that, though in some cases the responses for PS/PE and EOP are not for precise beats, these phrase parts can be identified by listeners in this study. As can be expected in such a free-choice, online study, there are some responses outside of these areas of majority preference. These may be due either to choice, or to 'mistakes' (chapters 4 and 10).

Overall, the profiles indicate that:
1) There seems to be a consistency between listeners, and
2) There is a systematic relationship between the three phrase-parts recorded. For each PS there is an EOP and PE.

Two types of relationship between PE and PS can be seen.
1) PE precedes PS (by up to half a bar) (Bach Suite, Bach Passion, Mozart Sonata).
2) PE and PS occur in the same beat (Mozart Aria, first PE of Bach Suite).

Similarly, two types of relationship between the EOP and the PE/PS can be seen:
1) EOP begins several notes before the PE (Bach Passion, later PEs in Bach Suite, Mozart Sonata, Brahms, Wagner). The amount of time before identified PEs varies from a couple of beats (e.g. Brahms) to a bar and a half (e.g. Mozart Sonata).
2) EOP is at the same position as the PE (e.g. Mozart Aria, the first PE in Bach Suite).

In general, when the PE precedes the PS, then the EOP precedes the PS as well. If the PE and PS peaks occur on the same beat, then the so does the EOP peak. The patterns in responses for PS/EP and EOP indicate that there is a consistency in understanding of the tasks.[30]

[30] Responses were also analysed on detailed timelines. Accuracy seemed independent of training or number of listenings (R.1.3, appendix 3.5).

In general, the areas of high response to the MIDI and performances coincide.

R. 2. Which sections of music are identified by listeners as phrases and what are their characteristics?

R.2.1 Length of sections

Phrases have been described as occurring within the 'perceptual present' and so should last between 5 and 9 seconds (for example, Snyder, 2000, see chapter 1).

Table 2.1: Average length of phrases, MIDI (PS to PS)

Piece	Average length (sec)	Standard deviation (sec)
Bach Suite	6.5	3.42
Bach Passion	8.16	3.45
Mozart Aria	5.6	2.86
Mozart Sonata	6.75	2.62
Brahms	5.7	4.32
Wagner	13.38	7.25

These results indicate that though the phrases indicated by listeners sometimes fall within this window, there are some that are longer.

R.2.2 To what extent are within subject and between subject results consistent?

There are many aspects of consistency and ways of measuring it (chapter 4). Here the general measure of the mean and standard deviation of PS are presented.

Table 2.2: Mean and standard deviation of PS presses per listener per piece

	Bach Passion	Bach Suite	Mozart Sonata	Mozart Aria	Brahms	Wagner
Mean PS presses per person	8.23	5.25	3.97	8.59	7.02	8.23
Mean Standard Deviation	1.75	0.78	0.45	2.75	1.54	3.83

The mean standard deviation of PS presses per listener per extract is small suggesting at least some systematicity in the responses. For the pieces for which the standard deviation is larger, there are musical reasons that lead to distinctly different phrase possibilities (chapter 10). The length of the excerpts varies.

In the following sections, for a general comparison of responses the proportions of key-presses in high-response areas identified from the profiles (appendix 3.6) were calculated; the number of presses in high response areas are divided by total number of presses (shown as percent). For some pieces, such as the Mozart Sonata, there are very clear areas, while for others, such as the Bach Passion, they are less clear. Other methods of analysis are employed in the following chapters (chapter 4 and 10). As the EOP responses are over a larger area, these responses are presented as graphs (appendix 3.6).

R.2.3 Comparisons between the pieces

The following tables show listeners' PS/PE responses for MIDI and performances.

<table>
<tr><th colspan="8">Table 2.3: proportion of responses within areas of high response, MIDI PS and PE (in %)</th></tr>
<tr><td></td><td>Bach Passion</td><td>Bach Suite</td><td>Mozart Sonata</td><td colspan="2">Mozart Aria*</td><td>Brahms</td><td>Wagner</td></tr>
<tr><td>PS</td><td>88.2</td><td>68.2</td><td>89.2</td><td>60.3</td><td>97.3</td><td>49.7</td><td>80.1</td></tr>
<tr><td>PE</td><td>83.8</td><td>96.5</td><td>90.1</td><td>62.4</td><td>91.0</td><td>40.6</td><td>57.8</td></tr>
<tr><td colspan="8">Comparison between high-response areas of different pieces</td></tr>
<tr><td colspan="8">The graphs (appendix 3.6) and table 2.3 show that the proportion of presses in high response areas differs among pieces. In general, the Bach Passion and Suite, Mozart Sonata and Aria have a large proportion of presses within high response areas, while the Wagner and Brahms have a lower proportion reflecting the difficulty of the task depending on the piece and its musical characteristics (chapters 10 & 11).</td></tr>
<tr><td colspan="8">Relationship between PS and PE</td></tr>
<tr><td colspan="8">The relationship between the PSs and PEs differs among the pieces. In general, the Mozart Sonata, Aria and Wagner have similar proportions of presses within high-response areas indicating that the PSs and PEs are equally clear. However, the Bach Passion, Suite and the Brahms have different proportions of presses within the high-response areas. For the Brahms, the proportion for PS is higher than for PEs, indicating that the PS is clearer than the PE. Conversely, for the Bach Passion and Suite, the proportion for PEs are higher than for PSs. Musical reasons for these differences are discussed in chapter 10.</td></tr>
<tr><td colspan="8">* The first number for the Mozart Aria excludes the 8 areas with smaller responses and the seconds includes them.</td></tr>
</table>

Comparing tables 2.3 and 2.4, (appendix 3.7 and the graphs, appendix 3.6) showing the average difficulty ratings for the PS/PE tasks for the MIDI renditions indicates that the latter correspond to the proportions of presses in high response areas.

Table 2.5.1: proportion of responses within areas of high response, Performances PS (in %) **

Sess-ion	Bach Passion		Bach Suite		Mozart Sonata		Mozart Aria*			Brahms		Wagner	
I*	F	96.3	G	85.8	U	86.4	S	83.0	83.6	K	51.1	Ba	58.2
II	F	94.3	G	80.8	U	93.3	S	66.5	84.5	K	63.6	Ba	48.5
I	L	94.4	R	75.8	L	97.5	B	59.2	87.8	Go	64.4	D	55.6
II	L	96.9	R	81.7	L	87.5	B	81.5	90.1	Go	63.8	D	51.6

Table 2.5.2 showing proportion of responses within areas of high response, Performances PE (in %)

Sess-ion	Bach Passion		Bach Suite		Mozart Sonata		Mozart Aria			Brahms		Wagner	
I	F	92.8	G	98.5	U	90.6	S	75.9	92.7	K	71.9	Ba	70.1
II	F	90.8	G	97.2	U	88.6	S	53.7	88.6	K	79.9	Ba	60.0
I	L	90.0	R	92.2	L	89.6	B	55.7	70.5	Go	91.8	D	49.8
II	L	91.6	R	99.5	L	83.8	B	82.4	72.2	Go	82.3	D	64.6

Comparison between high-response areas for performances of the pieces

The Bach pieces and Mozart Sonata have a generally high proportion of presses in high response areas, and small differences between responses for PEs and PSs and across the sessions. The Mozart Aria, Brahms and Wagner have a larger range of responses for PEs and PSs and across sessions

Relationship between PS and PE for performances

The relationship between PS and PE differs among the pieces. There is a higher proportion of presses for PE than for PS for all performances of Bach Suite, Brahms and Wagner. For the Bach Passion, Mozart Sonata and Aria, the situation is reversed with the higher proportion being for PS. The degree of difference between PS and PE differs across pieces but not greatly between renditions of the same piece. This is different to the responses to the MIDI where the PE are higher only the for the Bach Suite, the PS are higher for the Brahms, Wagner and Bach Passion and the PE and PS are the same for both of the Mozart pieces.

* Roman numerals refer to the session in which the recording was heard: I - first, II - second.

** F = Furtwängler, L = Leonhardt, G = Gendron, R = Rostropovich, U = Uchida, L = Lipatti, S = Solti, B = Böhm, K = Kovacevich, Go = Gould, Ba = Barenboim, D = De Waart

The numeric comparison above is supported by a statistical analysis of the differences. Here the whole response is considered.

Table 2.5.3: comparison between high response areas for the different performances and different sessions. Statistical test: Mann-Whitney[31]				
Piece	**Comparing responses from piece session**	**z-value**	**p-value**	**Difference significant?**
Bach Passion	Furtwängler I, Leonhardt I	0.87	$p \geq 0.05$	No
	Furtwängler II, Leonhardt II	1.09	$p \geq 0.05$	No
Bach Suite	Gendron I, Rostropovich I	2.40	$p < 0.05$	Significant
	Gendron II, Rostropovich II	0.26	$p \geq 0.05$	No
Mozart Sonata	Uchida I, Lipatti I	1.84	$p \geq 0.05$	No
	Uchida II, Lipatti II	1.59	$p \geq 0.05$	No
Mozart Aria	Böhm I, Solti I	1.73	$p \geq 0.05$	No
	Böhm II, Solti II	2.39	$p < 0.05$	Significant
Brahms	Gould I, Kovacevich I	2.64	$p < 0.01$	Significant
	Gould II, Kovacevich II	0.03	$p \geq 0.05$	No
Wagner	Barenboim I, De Waart I	4.32	$p < 0.001$	Highly
	Barenboim II, De Waart II	0.61	$p \geq 0.05$	No
This table indicates that there are three groups of pieces: 1) No significant difference between responses to the two performances (Bach Passion, Mozart Sonata), 2) A significant difference between responses to the recordings in the first session but not in the second (Bach Suite, Brahms and Wagner), 3) No significant difference in the first session, but a difference in the second (Mozart Aria).				

Table 2.6, appendix 3.7 shows the average difficulty ratings given by listeners to the PS/PE and EOP tasks in response to the performances. Again in comparison with the graphs and tables 2.5.1-2, there is a relation between the perceived difficulty and the proportion of responses falling in high-response areas. The first graphs in appendices 3.6.1-6 show different relations between EOP and PE/PS: 1) those that are prepared ahead of time and expected (such as in the Bach Suite), 2) those that are on the resolution note itself (such as the Mozart Aria), and 3) those that are last minute before the new start, that do not have a complete resolution (such as the Brahms). As with the PS/PE task, there are some pieces for which the expectation areas are clearer (such as the Mozart Sonata) and some for which there are more key presses between the peak areas (such as the Bach Passion).

R.2.4 Summary

This general comparison of the responses to the different pieces indicates that:

1. There are different proportions of responses in areas of high-response in the different pieces.

---

[31] Many thanks to Dr. Pentecost for his advice on methods of statistical analysis of the data.

2. Areas of response for the same piece in different renditions stay constant.
3. The proportions in these areas in some cases differ and in others do not between renditions.
4. The relation between EOP, PS and PE responses changes among pieces
5. There is a general relationship between perceived difficulty and key-presses in areas of high response.

R.3. Are there differences between the responses related to musical experience?

The results are analysed by comparing two and three subgroups of listeners according to musical education. The results for the two-group comparison are in appendix 3.7 (R 3.1). For the latter, the listeners were divided into: DL, M and N (section 3.3.2.3).[32]

R.3.1 PS and PE MIDI and Performances

The comparisons in section R. 3.1 appendix 3.7 show that there are differences between the three groups but they are not consistent in proportion or direction. For the MIDI PS for most of the pieces the differences are relatively small between DL and M. The results for the Ns though lower than for the other groups in half of the pieces (the Bach Passion, the Mozart Sonata and the Brahms), show a relatively high agreement for most pieces. The difference in responses for the Wagner between DL and M is unusual both in that the difference between the two groups is larger here than for any other of the pieces and that the difference is the reverse (response is higher for the M than for the DL).

For the PS and PEs for MIDI and performances within the groups there are variations (sometimes larger than between groups); the responses are not consistent across pieces, and none change systematically according to education. In summary, these results indicate that for most of the pieces, the different ability groups (and especially the DL and M) may, for the most part, be treated together.

R.3.2 EOP for MIDI and Performance responses

The comparisons for EOP responses for DL, M and Ns in (section R. 3.2 appendix 3.7) show that, in general, the responses of the three groups occur in the same areas. The relationships between the three groups change within and among the pieces, sometimes a group preceding or being spread over a smaller area than the others. However, there is not a systematic difference between them.

[32] As there are only 5 Ms for the MIDI, statistical tests were not applied.

R.3.3 Summary of comparisons of responses from listeners with different training

For all of the training groups there seem to be a systematic relationship between EOP, PE and PS. There does not seem to be a clear consistent difference between the groups. There are some cases in which there are differences between the groups' responses (such as some of the EOP responses starting later). As will be discussed in chapter 10 some of these differences can be related to the musical features.

R.4 Does prior familiarity with the piece affect responses?

Listeners were asked whether or not they recognised the pieces and were here split into two groups according to this. There were not many listeners who recognised the music. The comparison was only made for pieces for which at least 5 listeners identified the piece. The responses were analysed in two ways: 1) using the same approach as above; comparing the responses within areas of high response and 2) comparing the responses in different areas of high response within the piece (for PSs).

R.4.1 Responses by those with and without prior familiarity with the piece

Table 4.1: proportion of responses within areas of high response for listeners with and without prior familiarity with the piece (in %) (MIDI)

	Bach Passion	Bach Suite	Mozart Sonata	Mozart Aria	Brahms	Wagner
With prior familiarity with the piece	86.1	64.1	94.2	-	61.3	-
No prior familiarity with the piece	82.4	69.5	88.0	-	47.3	-
Number of listeners with prior familiarity with piece	6	8	5	2	6	3

Comparison of responses of those with and without familiarity with the piece

Three of the four pieces have higher responses in the group familiar with the piece indicating that there may be an effect of familiarity. The difference is greatest for the Brahms, which may benefit most from familiarity with performance characteristics (chapter 7). A similar comparison was attempted for the responses to the performances. Observations indicated that the two groups' responses were similar. However, as the number of listeners who both recognised the pieces and had not taken part in the MIDI study were small (four or less for all the pieces), it was not possible to carry out the same comparison.

R.4.2 Does prior knowledge result in longer phrases?

The graphs in appendix 3.7 compare PS responses for the high-response 'phrase' areas within each piece of those 'familiar' and not familiar with the pieces. Graph 4.2.1, appendix 3.7 of the Bach Suite shows that there is almost no difference between the groups with prior knowledge of the piece and those without. There is a larger proportion of PS responses 3 and 4 than for any other regardless of familiarity. Graphs 4.2.3 and 4.2.5, appendix 3.7 of the Mozart Sonata and Brahms show that a smaller proportion of those who knew the piece prior to the session pressed a key in some phrase areas. More of those who knew the piece identified longer phrases than those who did not. Comparison of the group who were familiar with the piece with the rest of the DL and non-DL (graphs 4.2.4 and 4.2.6, appendix 3.7) show that those familiar with the piece have the strongest manifestation of these 'longer' phrases, but the DL and non-DL also show this trend (although the non-DL have a lower response for the first phrase).

R. 4.3 Summary of prior familiarity with the piece

Sections R. 4.1 and 4.2 indicate that familiarity with the piece before the session may affect phrase identification in some pieces; in three of the four pieces there is a larger proportion of presses in 'phrase' areas by those who were familiar with the piece than by those who were not. Furthermore, there may be a relationship between prior familiarity and phrase length in some pieces (and not simply levels of training).

R. 5. Are there differences between consecutive listenings by the same listener. If yes, what are they, and do they indicate learning?

R. 5. 1. Indications for Learning within the set of three listenings PS/PE

For the MIDI responses, tables 5.1.1 and 5.1.2 in appendix 3.7 show that there is little change in the proportion of responses in the high response areas from the first to the final listenings for the PSs. None have the pattern of gradual 'improvement' over the three listenings indicating that over the three hearings, there is no 'learning'. The PE show a similar pattern to that of the PS responses; none of the final listenings have the highest proportion of key presses in high response areas. Statistical tests of all the responses showed no significant differences. Moment-to-moment reaction seems to be winning over the long-term schema of what they decided to do following a complete listening.

Though listeners sometimes made retrospective comments verbally about wanting to do things differently in subsequent listenings, this does not seem to be implemented systematically.

For the performances, tables 5.1.3 and 5.1.4 in appendix 3.7 show that for the PSs there are six performances for which the highest proportion of presses in high-response areas are for the final listenings. Only four of these follow a 'systematic' improvement within sessions. For the PE there is almost no difference between the listenings: only the two performances have a gradual 'improvement' over the three listenings.

R.5.2 Indications of learning within the set of three listenings, EOP

Section R.5.2 in appendix 3.7 indicates that again there is no systematic change in responses across the listenings. For some pieces, the responses of the last listening are earlier than the previous ones and overall, the responses are 'tidier' in the third listening than in previous ones for some positions (Bach Passion MIDI). For others the responses move closer to the PE (for example, Bach Suite, bar 2). For others (such as, Bach Suite, bar 3) there is no change, the position seems clear from the first. For several pieces, there is little change across listenings (such as the Bach Suite performances and the Mozart Sonata and Aria MIDI and performances and Wagner). For others, there are changes, though they are not systematic (Brahms).

Overall, there does not seem to be a systematic trend in direction of change of peaks for the EOP responses related to learning.

R.5.3 Indications of learning within the set of three listenings, Summary

Overall, there did not seem to be a systematic trend across tasks and pieces related to the number of times the piece had been heard; 'learning', if it occurs, does not seem to have a systematic effect (see also chapter 4).

R. 6. Are there variations between listeners' responses affected by the order of pieces or tasks carried out and if so what are the implications? [33]

R. 6.1. Variation as a result of having task order?

Table 6.1, appendix 3.7 shows that are only two pieces for which there is a relatively large difference between those who did the PS/PE or the EOP task first (Bach Passion and Mozart Sonata). In both, there is a large increase in the proportion of responses in areas of high response for those that did the EOP task first, indicating that there may be an effect of getting to know the piece or thinking about it in a different way before doing the PS/PE task. However, for these pieces there was also a relatively large difference in the other measures discussed. Moreover, for the other pieces there are very small changes in the other direction or no change at all.

The difference between the results for different pieces may be related to the relative ease of the tasks for each excerpt. The Mozart Sonata may be the clearest

[33] Similar comparisons were made for piece order; no systematic differences where found.

(with highest response when carrying out the PS/PE task first) and therefore more listenings may reinforce the decisions. On the other hand, it seems that if the decision is less certain in the excerpts that are more complex, the listeners firstly concentrate only on the task at hand and do not relate it to the broader context, and that more listenings or doing the other task first does not have an effect.

The results of R. 5. and R. 6 indicate that once presented with a MIDI version of a piece, the phrase identification does not change systematically with further listenings.

There may be a difference between responses to listenings that follow each other relatively quickly, the rate of which is designated by the experimenter, as was the case here, and listenings that occur at the rate the listener chooses. If the listeners had been given more time (or been able to choose when to begin listening again), their choices may have changed more. This study indicates that there are no consistent changes across listenings (both within task and task ordering). All the results are taken as one group.

R. 6.2 Effect of having heard the MIDI versions first on responses to performances

Fifteen listeners returned six months after they took part in the MIDI sessions for the performance sessions. Their performance responses were compared with those that had not heard the MIDI renditions. Table 6.2, appendix 3.7 shows that for the first session (I) in all but the De Waart and Kovacevich there are a higher proportion of key-presses in the high response areas for those that had heard the MIDI previously. However, the differences are, for the most part, very small. The differences are even smaller by session II. It should be borne in mind that the proportions of listeners with different musical ability are not always the same in this comparison.

3.3.4 Summary

S.1. What is the nature of the written, verbal and musical responses?

S.1.1.1 Written responses

The written responses indicate that the term ‘phrase’ was meaningful to listeners. The terms used have their origin in a number of disciplines and there does not seem to be an obvious link between the types of terms used and the musical experience of the listener (though there is sometimes concentration on some terms by one group). The general ideas associated with the term ‘phrase’ seem to be similar.

Some elements are not mentioned in the listeners' definitions, particularly in the MIDI responses. For example, the presence and description of the 'beginning' of a phrase is given only once, and even then it is in the context of the span of the whole phrase. This contrasts with the greater number of references to phrase ends and their characteristics (if only general). This bias may be because of the expectation task. However, the listeners were asked to identify phrase starts as well as phrase ends. It may therefore be that the bias is because either the start seems too obvious to mention or the listeners do indeed concentrate more on the PEs than the PSs.

Comparison between responses at the end of the MIDI and performance sessions indicates that the rendition type has little effect on the definitions of the term. The overwhelming majority of responses indicate that performance features do not form part of the written definition of the phrase given after performances, and are not "missed" in the MIDI renditions.

S.1.1.2 Verbal responses

Unlike the written definitions, the responses after each piece refer to musical characteristics of the particular piece and point out the differences between what listeners expected and the modifications that the listeners made to those expectations.

There may be a 'list' of ideas that listeners are using, as indicated by the frequent reference to 'missing' features. However, the listeners are flexible when listening: the terms used by the same listener vary between and within the excerpt (for example, listeners comment on the change of usefulness of different characteristics within the excerpt).

Unlike for the written definitions, in the verbal responses after each MIDI rendition, there seems to be a broad distinction between a relatively small number of categories chosen by DLs and non-DLs. In contrast, after each performance rendition, the verbal responses seem to be from a greater variety of categories with less systematic distinction between DLs and non-DLs and in one case there even seem to be different categories depending on the rendition.

Both DLs and non-DLs comment equally on difficulties encountered when responding to the performances though sometimes the difficulties are described differently. One listener comments on the lack of performance features in the MIDI. Some listeners even mention 'performance features', such as getting quieter or slowing down during the MIDI session. There is a rather large reference by listeners to performance features. However, there is also a surprisingly large proportion of listeners who find the performance features contradict other 'musical' features or make it more difficult to identify phrases. Moreover, when a contradiction between performance and 'other' features is mentioned, the listeners tended to prefer the other features to the performance ones. This indicates that these listeners have an idea of the phrase structure that is separate from the one presented by the performer, indicating that when listening to performances at least some listeners are not purely reliant on the performer's features but have their own interpretation and expectations based on other musical features. Those who mentioned the performance features in this way included listeners who had not heard the MIDI renditions before i.e. had not heard the pieces without performance features. Indeed, several of these listeners then made value judgements about the performances in general (usually negative). It seems that these features that affect phrase perception may also be influencing the extent to which we like a particular performance.

S.1.2 Comparison with the literature

Comparison between the written and verbal responses obtained in this study and the terms used in the literature shows that there is an overlap and that each study concentrates only on a small number of the aspects listed. Many of the terms used in the literature and in the responses obtained in this study are broad or even unclear.

One of the differences between the discussions in the literature and the responses given here is that the nature of the boundaries between phrases is not developed in these listeners' definitions. Though several musical features are mentioned, none (except harmony) are described in terms of how they are used to identify phrases or their boundaries. There are some musical features that are often mentioned in the literature, such as repetition, that are not mentioned here.

S.1.3. Timeline, high response areas, and the nature of the responses to MIDI and performance

Many questions are raised when presenting and analysing this form of raw, online, free-choice data. Most previous phrase research avoided such questions by setting up experiments that elicit different kinds of responses or similar responses that can be prepared and therefore have less variation. Here, tactus units were used in a time line representation. Peak areas were identified and taken as broad response areas. To allow comparison between responses to different pieces or renditions and responses from different groups of listeners, the proportion of responses within all the peaks were here represented together numerically allowing a broad description of the data. This establishes a basis for comparison and further analysis.

The results indicate a consistency in understanding and response to the phrasing and expectation tasks with different relationships between PS, PE and EOP, and size of areas of peaks, both depending on the excerpt.

S.2. Which sections of music are identified by listeners as phrases and what are their characteristics?

The responses indicate that the lengths of phrases chosen are mainly within the range of 5-8 seconds with one piece having an average phrase-length of 13 seconds. The standard deviation varies from 2.6-4.25 with one piece having a standard deviation of 7.25. The units are here sometimes longer than those often described as phrases by music psychologists of 2-5 seconds (with extremes of 12, chapter 1). There seems to be a range of phrase lengths depending to some extent on the piece's characteristics.

The positions of PS, PE and EOP of the phrases relative to each other, and the duration of the period of high responses vary among a number of options in different musical contexts.

The proportion presses in high response areas and the relative proportion for PSs as compared with PEs differs among and within pieces. In some cases, the proportion of key presses is similar, indicating a similar level of ease/difficulty in identification. In others, the proportion is higher for one than the other, indicating a difference in difficulty between PSs and PEs.

S.3. Are there differences between the responses related to musical experience?

Overall, the MIDI results show that the response areas are similar for DLs and non-DLs, often with slightly higher responses for DLs than non-DLs. However, the relationship changes from between and within pieces. The performance responses indicate that for PSs the responses from DLs and non-DLs are generally similar, while for PEs, there are some differences. For some pieces, the relationship between DLs and non-DLs stays the same for the different sessions, while for others it changes.

Comparison of the DL, M and N shows that there are differences between the groups but they are not consistent for all pieces in proportion or direction; which group has a higher response. The N's responses are lower than the other groups in only half of the pieces. In some cases, the N's EOP responses are later than those of DLs and Ms.

In general, the difference between DLs and Ms is small. Moreover, within groups there are variations that are sometimes larger than between-group differences. It seems, therefore, that for most of the pieces, all three ability groups may be treated together. As will be shown, the differences in responses may be explained by musical characteristics rather than only by the musical experience of the listener.

S.4. Does prior familiarity with the piece affect responses?

There were higher proportions of responses and longer phrases for some pieces.

S.5. Are there differences between consecutive listenings by the same listener. If yes, what are they, and do they indicate learning?

Overall, there is no systematic pattern, such as convergence to the high-response areas, over the three listenings. There is a systematic increase over the three listenings in only two performances. If there is a process of learning over the three listenings, it is not represented in the measures investigated here.

S.6. Are there variations between listeners' responses as a result of the order of pieces or tasks carried out and if so what are the implications?

In the MIDI study, for the majority of the pieces the response rate in the high-response areas for the PS/PE task is not increased by having carried out the EOP task. It seems, therefore, that not only is there little effect of short term learning between attempts, but there is also little affect of more familiarity or carrying out the other task on the same piece. These findings contradict the impressions of listeners who, in verbal communication during the sessions, said that they found the PS/PE task easier if it was preceded by the EOP task.

In the second performance sessions, listeners were asked whether or not they recognised the pieces from the previous session. The majority did and for most of the performances (except for the, Mozart Aria and Sonata) commented that the performances were different. There seems to have been a long-term memory effect (two-weeks) of having heard one performance on responses in session two.

In general, these results indicate that though there are some differences in responses between listeners, none of the above variables (musical experience, prior familiarity with the music, learning within and from an experimental session) account systematically for the differences observed. There may be others ranging from specific preferences of listeners to one genre of music, to the position of the piece in the experiment, to how tired the listener may be to whether they are male or female (Baron-Cohen made explicit reference to, on average, males' superior ability in structure identification in music in 2005a; Baron-Cohen, 2005b). In terms of this last, responses from males and females of this small sample were compared and no significant difference was found.

Therefore, a method was developed to identify possible groupings of responses beginning with the data. This begins with a quantification of listeners' self-consistency, identifies subgroups and quantifies their between-listener consistency and shows that there are a number of clear candidate positions that can be chosen in different combinations (chapter 4). This paves the way for the following chapters which explore possible musical reasons for the responses.

Chapter 4

Listeners' Phrasing Study – A statistical method: Analysis of listener's self-consistency and population agreement

An after-phrase

4.1. Introduction

The previous chapter analysed the listeners' responses to the tasks of identification of phrase starts (PSs), phrase ends (PEs) and the beginning of the expectation of the end (EOP) primarily starting from the general view of identifying locations of these responses and differences between groups of listeners from the perspective of the whole data set. All listeners' three listening responses were included and compared. Here, the listeners' responses are analysed from the opposite direction: from the individual to the general.

This method, using the kappa (k coefficient) statistical test, was developed originally for the analysis of responses in language annotation and perception experiments in which responses are to a classification task (such as Krippendorff, 1980). The main aim of the tests is usually to assess the degree of consistency of individual responses and population agreement, i.e. assessing the level of repetition, and is often used as a test for the possibility of a (conscious or unconscious) 'rule base' (such as Carletta et al., 1997).

In this chapter, listeners' responses to the phrasing task discussed in chapter 3 are analysed in order to investigate whether or not the inter- and intra-listener responses are 'systematic' i.e. do they have similar responses to the same piece a) during repeated listenings, and b) to each other. Here, unlike in other studies that use a similar approach (Krippendorff, 1980) the search is not for complete agreement; listeners were explicitly encouraged to change their minds. Instead, this method is used to begin to distinguish between different possible types of differences between responses both for the same and between listeners.

Much of the method used, particularly in the way the 'interpretations' are reached, was developed here. The method has several stages that depend on one another. Therefore, the method section includes interim results used for further analysis in subsequent sections.

4.2 Aims

The aims of this method are:

1. To determine the level of consistency of decisions made by the same listener during three consecutive listenings to the same rendition.
2. If individual listeners are consistent, to represent their phrase responses as a single response for each rendition which is then taken as their 'interpretation' of the phrase structure of this rendition.
3. To determine the level of agreement between subjects about the phrase structure of each piece and investigate the possibility of sub-groups of agreement.
4. To compare responses to the PS, PE and EOP and to different renditions of the same piece.

This data was collected continuously (listeners could press at any moment while listening) and is represented using timelines with units that can vary in nature and magnitude. Two preliminary questions are pertinent to the analysis of this kind of data and answers to these determine the method of application of the kappa statistical test.

a) Which basic unit?
The data can be presented in relation to real time (for example as equal length units in seconds) or in relation to a musical unit: the note length, beat, bar etc. In other words, what is the size and definition of the 'unit' for which PS, PE or EOP responses within that unit can be considered 'the same'? Here, several options are explored.
b) In what way can the 'areas' of response be represented?
Once a unit has been chosen, as the graphs in each of appendices 3.6.1-6 showed, there are often 'areas' of response and these differ in size (depending on the musical features, chapters 10-15). Here, the identification and definition of those areas are explored in a systematic numerical manner.

4.3 Theory

The method identifies all the possible positions of responses and compares the probability of a non-response with the actual non-response. '[T]he κ coefficient of agreement has become the de facto standard to evaluate intercoder agreement for tagging tasks' (Di Eugenio and Glass, 2004, p. 95). It quantifies the degree to which agreement is not due to chance (denominator of equation 1) i.e. the non-random component of the observed agreement between listeners (numerator of equation 1)[34] (Siegel and Castellan, 1988).

> 'κ is computed as P(A)-P(E)/1-P(E) , where P(A) is the observed agreement among the coders, and P(E) is the expected agreement, that is, P(E) represents the probability that the coders agree by chance. The values of κ are constrained to the interval [-1, 1]. A κ value of one means perfect agreement, a κ value of zero means that agreement is equal to chance, and a κ value of negative one means "perfect" disagreement' (Di Eugenio and Glass, 2004, p. 95).[35]

Let N be the number of items to be classified; m - the number of categories to classify into (here, $m = 2$); k - the number of listeners; n_{ij} is the number of listeners who assigned the i-th item to j-th category.

$$\kappa = \frac{P(Actual) - P(Expected)}{1 - P(Expected)} \tag{1}$$

$$P(Expected) = \sum_{j=1}^{m} p_j^2, \quad p_j = \frac{\sum_{i=1}^{N} n_{ij}}{Nk} \tag{2}$$

$$P(Actual) = \frac{1}{Nk(k-1)} \sum_{i=1}^{N} \sum_{j=1}^{m} n_{ij}(n_{ij} - 1) \tag{3}$$

[34] Thanks to Prof. Eli Shamir for providing this explanation and Beata Beigman Klebanov for discussions about, and implementation of the method.
[35] In the formlae, Actual = A and Expected = E.

This statistic is employed in content analysis (Krippendorff, 1980) and in computational linguistics where human annotators are used to create gold-standard data sets for classification tasks, including reference-type classification of definite descriptions (Poesio and Vierira, 1998) and dialogue classification (Carletta et al., 1997).

In general, agreement levels of $\kappa \geq 0.67$ are considered high enough to tentatively conclude that subjects are working under the same understanding of the phenomenon. The threshold is based on extensive studies and is a useful guideline (Carletta, 1996; Carletta et al., 1997; Di Eugenio and Glass, 2004; Krippendorff, 1980). If these agreement levels are obtained, the data set is considered reliable enough for testing computational models of the phenomenon.

However, difficulties have been identified. For example, slightly different methods of calculating the same statistic, based on different assumptions, can lead to slightly different results (Di Eugenio and Glass, 2004, p. 95). The value of 0.67 adopted as a cut-off in computational linguistics above which there is agreement (Di Eugenio and Glass, 2004, p. 95) is based on the assessment of κ values in Krippendorff (1980), which discounts $\kappa < 0.67$, allows tentative conclusions when $0.8 \geq \kappa \geq 0.67$ and definite conclusions when $\kappa \geq 0.8$. However, some, including Carletta et al. (1997) and Di Eugenio (2000), have warned against using this as a standard. For example, for Rietveld and Hout (1993) $0.40 \geq \kappa > 0.20$ indicates fair agreement, and $0.60 \geq \kappa > 0.40$ indicates moderate agreement (Di Eugenio and Glass, 2004, p. 97).

Here, the κ statistic is used to quantify the degree of each listener's self-consistency across trials and the agreement between listeners.

4.4 Method of analysis

The method is discussed first in relation to the MIDI PS responses. The discussion of the rest of the MIDI data follows. The results of the performance data are presented and compared with the MIDI results in section 4.5.

4.4.1 Calculation of individual's initial self-consistency and identifying listener's 'interpretation'

The quantification of the listener's self-consistency includes an 'interpretation', representative of their three responses, and a numerical assignment of the agreement according to a statistical cut-off.

a) The representation of an 'interpretation'
There are several ways of representing each listener's 'interpretation' (Deliège, 1998 and chapters 1 and 8). Here, the positions identified in the majority of the listenings are taken as part of the interpretation (i.e. if a position was pressed at least twice out of the three listenings, except the Mozart Aria which had many

cases of only two listenings because of time constraints). If a position was chosen only once it may be a 'mistake' or an option that is less obvious or important to the listener (these presses are discussed in chapters 3, 10 and 11). This method incorporates the idea that listeners were not asked to be consistent or to remember their previous responses; they could change their minds or provide different options. It identifies the positions that the listener found most 'important' (as represented by their repeated choices).

There are several ways of quantifying the self-consistency on the basis of the 'interpretation'. As the common positions from the three listenings are already represented in the 'interpretation', here the 'interpretation' is compared to the listening that is most similar to it, thus comparing the listening with the least 'mistakes' or one-off presses with the 'interpretation'.

b) The minimum number signifying statistical agreement
The numeric threshold for the kappa statistic varies among studies but is often taken as 0.67. Here there could be some dependency between responses; although the listeners were told they could change their minds, they are still the same listeners. Therefore, for self-consistency the threshold is increased to 0.7.

4.4.2 Timelines and areas, close-to-note-boundary presses, unit merging.

Listeners could press a key at any moment while listening. The beat, level of detail or subdivision identified by the listener may depend on, for example, the musical context and structure (controlled by the composer), the structure of Western tonal music, and expectations and 'attending' of the listener (Jones, 1992, pp. 91-95). There are, therefore, several options of how to represent the responses on a time-line (chapter 3). Here, various options for the subsequent analysis are considered: equal units of real-time, note length, the beat and their combinations.

Real-time units, for example seconds, can be applied to all pieces in order to allow direct comparison. This measure also helps to clarify the contention that phrase length is limited to a certain real-time duration (7-9 seconds) because of our memory and other psychological limitations. However, studies have shown that the perception of time changes according to different tempi of music and the level of events per unit time has an affect on its perception. Moreover, real-time does not seem to play a large part in our temporal perception of music. Listeners are not good at time perception. Indeed, listeners' perception of time, seems to be related to the events per unit time, rather than absolute length (chapter 5.5.3).

Taking the individual notes of the piece as the timeline units relates directly to 'what is played'. Whilst this seems simple for monophonic pieces, there are several options in polyphonic ones: should the melody line be used, the melody and the bass, or any part that sounds at that moment. Moreover, in both monophonic and polyphonic pieces there are cases in which there are many fast notes. In these cases, pressing on a specific note is very difficult. Experiments have shown that the limit of distinction of individual notes is estimated to be 2ms while at least 100ms between note onsets are necessary to identify more information and beats are perceived in a range of 200ms to 1.5-2 seconds (many such studies include Bolton, 1894; Hirsch, 1959; London, 2004).

Using a regular beat (usually the tactus, chapter 1) for PS/PE responses to some extent overcomes this problem; the short notes that are difficult to pick out are considered as part of larger, musically meaningful units. This is the method used here. If the regular beat is taken, there may be beats on positions in which there is no note onset. In such cases, the nearest note-onset is taken. As the EOP can have a larger area of response than the PS/PE, these results are also presented on 'area' timelines combining the 'areas' of response seen on the crotchet timelines (appendix 3.6).

4.4.2.3 Self-consistency figures

The self-consistency figures were calculated using the responses organised on the timelines, and the comparison between the 'interpretation' and the most similar listening using the kappa statistic (table 4.4.2.3, appendix 4). Graph 4.4.2.3.1, appendix 4 shows the number of listeners with self-consistency figures in three categories: complete agreement, agreement, below agreement threshold. In all but the Mozart Aria the number of listeners with consistent responses is higher than the number whose responses are not. The agreement figure can fall below 0.7 for two reasons: 1) the presses are not in exactly the same position but are close to each other (graph 4.4.2.3.2, appendix 4) or 2) the presses are far apart (graph 4.4.2.3.3, appendix 4).

A strategy can be developed for identifying presses that are very close to each other (as in (1) above), and also explore the idea of PS, PE and EOP 'areas' rather than 'positions'. There are often presses that are very close to the timeline's unit-boundaries. In some listenings, presses are close to the end of one unit and in others they are close to the start of the next. In several of these cases, there are also presses well inside both of the units. This distribution of presses indicates that: 1) either there is anticipation or late response to particular notes and/or 2) they may be considered, at one level at least, equivalent for phrase perception. These units are therefore 'merged' helping to identify the 'equivalent' presses and 'areas' of PS, PE or EOP.

In this implementation, for equivalent close-to-boundary-presses there have to be at least two presses at the boundary, at least one person has to have pressed on both sides of the unit boundary, and they have to be within a 2 ms range.

4.4.3 Recalculation of individuals' self-consistency and 'interpretations' using timelines with merged units

The individuals' self-consistency and 'interpretation' were re-calculated using the unit-merging according close-to-boundary presses.

This merging results in changes in some of the consistency figures. There is a substantial improvement, indicating that much of the 'inconsistency' identified in the original timelines originated from close-to-boundary presses i.e. is not considered inconsistent in this method. For example, graph 4.4.3.1.1, appendix 4 shows the responses of the same N as above (section 4.4.2.3, graph 4.4.2.3.2, appendix 4) after unit-merging. In contrast, there are cases, in which some responses by the same listener are far apart for different listenings, for which the unit-merging makes little difference to the consistency (k stays below 0.7). For example, graph 4.4.3.1.2, appendix 4 shows the listener responses for the DL listener whose results were discussed above (graph 4.4.2.3.3, appendix 4).

For most individual listeners there is a mixture of close-to-boundary presses and more distant presses. At this stage it is not possible to distinguish between a 'mistake' and a less favoured but valid option. However, as will be discussed below (section 4.5.4.2), once individuals' responses are compared with one another, this can become clearer.

The results are discussed from the perspectives of the piece and the listener.

4.4.3.1 By piece

The difference between the number of listeners whose results are above the consistency threshold before and after unit-merging for the individual pieces is summarised in the table 4.3.3.1.

Table 4.3.3.1: Effect of merging units in three groups, by piece, for MIDI, PS

Piece	k is lower in raw than merged	k is lower in merged than raw	k < 0.7 in both raw and merged
Bach Passion	6	1	3
Bach Suite	5	0	0
Mozart Sonata	3	0	2
Mozart Aria	7	0	5
Brahms	5	0	2
Wagner	0	0	0
Total	26	1	12

In total there are 26 cases below the 0.7 threshold before merging but above it after merging. There are only 12 that remain below the threshold and there is only one that was not below the threshold before merging but now is (see table 4.4.2.3, appendix 4 for the individual results). The 26 cases that have lower self-consistency figures in the raw than merged results were initially 'inconsistent' because listeners were pressing in areas around boundaries but not exactly on the same note. The 12 that remain lower than the 0.7 threshold after unit-merging, are presses that fall outside these areas, (usually) in completely different positions – suggesting different alternative interpretations. This does not happen at all in the Wagner: all the raw and merged fall above the 0.7 threshold. In the Bach Suite, there are only examples of the first category. The Mozart Sonata and Brahms have improvement after unit merging. The Mozart Aria has the highest number of examples of the second category. In some of these cases, at least some of the positions chosen are very far from one another (there is one 'missing' or 'additional' position). In others, the positions in the different listenings are close but not close enough to fall within the same unit or merged areas.

There is only one example of κ now being below the 0.7 threshold when the raw was above (in the Bach Passion). Although the areas identified are the same, there is a larger area of confusion for this listener (a N). As the method uses the total number of units to calculate the consistency figures, in this case the reduction in the number of units (after merging) has a detrimental effect on the self-consistency figures.

4.4.3.2 By listener

Table 4.4.3.2 shows the number of listeners with self-consistency in four categories, after unit merging.

Table 4.4.3.2: Effect of merging units with close-to-boundary presses in four groups, for MIDI, PS (out of 34 listeners)

	k = 1 for all pieces	k ≥ 0.7 for all pieces	k < 0.7 for 1 piece	k < 0.7 for two pieces	k < 0.7 more than 2 pieces
Listeners	1	22	10	2 (Ns)	0

These results indicate that: 1) there are no listeners that have consistently inconsistent responses and 2) most listeners are self-consistent in most pieces.

4.4.4 Calculation of inter-listener agreement of 'interpretations' and identification of groups of 'interpretations'

4.4.4.1 Agreement between interpretations of all listeners for each piece

Having established the timelines and identified individuals' 'interpretations', it is possible to compare between listeners. The level of agreement or consistency between listeners is calculated in the same way as for the self-consistency. The 'interpretations' are compared with each other in pairs. The comparison is between independent responses and so the threshold for pairs of inter-subject consistency is κ = 0.67 (section 4.3).

Table 4.4.4.1: Agreement for the whole group per piece, MIDI, PS						
	Bach Passion	Bach Suite	Mozart Sonata	Mozart Aria	Brahms	Wagner
kappa (k)	0.40	0.71	0.63	0.28	0.41	0.45
Z score	5.22	12.48	7.31	3.40	6.63	6.03
Z scores are all significant at $p<0.001$ level						

These figures are rather low though the Bach Suite (Mozart Sonata, and even the Bach Passion, Brahms and Wagner according to Di Eugenio and Glass, 2004, section 4.3 above) show(s) group consistency at a 'significant' level. However, some subgroups of interpretations are more consistent.

4.4.4.2 Agreement within groups of listeners

Subgroups of interpretations are identified by 1) pairing interpretations that are most similar to one another and 2) building groups from the individual listener that is common to the most pairs. The positions that are most common in each group also occur together in at least one (usually more) individual. The groups of interpretations are discussed below (section 4.5.4.1.1). The positions chosen by the different groups are identified and compared here and will be discussed in relation to the musical features in chapter 10.

4.5 Results and Discussion

4.5.1 Initial individual's self-consistency and 'interpretation'

Most of the listeners have significantly self-consistent responses (section 4.4 and table 4.4.2.3, appendix 4). The inconsistent responses were classified into two types – those of presses in the same small area, and those that are more distant (sections 4.4.2 and 4.4.3). Merging the units of the time line according to confusion points (section 4.5.2) improves the self-consistency figures for many of the cases that fell below the consistency threshold. The results, therefore, allow the distinction between the two types of 'inconsistency' discussed above (sections 4.4.2 and 4.4.3) and sets the basis for comparison between listeners' interpretations (sections 4.5.3 and 4.5.4).

4.5.2 Timelines and areas. Close-to-note-boundary presses, unit merging.

4.5.2.1 The timelines

For the PS and PE, musically, technically and psychologically the most suitable timeline seems to be one with beats as the basic unit. If there is no note-onset on the beat, the nearest following note-onset is taken. As the EOP can have a larger area of response, these results are also presented on 'area' timelines (section 4.4.2).

4.5.2.2 Merged areas

Each piece had enough close-to-note-boundary presses to have merged areas the amount of which varies between pieces (graphs 4.5.2.2, appendix 4). The majority of positions included in the final interpretations have some degree of merging. This could be interpreted as an indication that it is only possible to find agreement when using relatively large units. However, it could also be that it is difficult to press on exactly the same note in repeated listenings, technically, musically and psychologically. For example, during online listening to some extent the location of the PS is dependent on the location of the previous PE, and sometimes listeners press the PS while still on the PE note in early listenings and later in later ones. Moreover, the degree of unit merging varies depending on the musical context.

Furthermore, it is not merely a case of the larger the area merged, the higher the agreement. For example, in the Bach Suite, a small range (2-2.375) is merged near the start of the piece; the amount of confusion is limited to a small area. Musically, there may be two possibilities: the PE and PS may coincide on 2, or the PE may be on 2 and the PS on 2.375 (chapter 6). Another small range (6.875-7) is merged. Here the response on 6.875 seems to be an anticipation (sometimes very close) of 7 – the beginning of the PE. The long chains of 2.875-4.5 and 5-6.5 seem to combine two aspects. 2.875-3 seems to be equivalent to 6.875-7, and 3-3.25 and 5-5.25 are the PEs and 3.5 and 5.5 next PSs. It is not surprising that there are presses throughout this area. From 3.875-4.5 and 5.875-6.5 there are further features that could indicate PSs (chapters 10 and 11). These indications are responded to by listeners and, because of their proximity to the PSs, are merged with them. These results indicate that there is a relatively large area in which there are PS responses indicating a large area including stronger or weaker musical cues. The situation is clearer for the PE for which the merged areas are much shorter and distinct (2.875-3.5, 5-5.5, and 5.875-6.375) indicating that the PEs areas seem clearer than the PSs.

In general, this analysis of individual and between listeners' consistency indicates that, in some cases, the listener responses are spread over areas. This may be because of short-term delay or anticipation (such as in 6.875-7 in the Bach Suite) or inaccuracy but often implies longer-term anticipation or delay in response to musical features (chapters 10-12). Boundary areas (rather than positions) would be concealed if we looked only at specific notes as positions of phrase boundaries (chapter 8). The merging procedure therefore highlights the boundary areas.

4.5.2.3 Comparison of merged areas

Comparison of merged areas in responses to the MIDI and performance sessions shows that, in general, the merged units occur in the same areas but the areas are sometimes smaller in the performances than in the MIDI. The similarity indicates a commonality in the responses to the two types of renditions and the smaller areas of response to the performance renditions indicate that here there seems to be a clearer preparation and accentuation of phrase boundary areas (graphs 4.5.2.2, appendix 4).

Both the MIDI and performance results show that the PE areas usually begin earlier, but often overlap with, PS areas. EOP areas tend to be the largest but also overlap, or are close to, PS and PE merged areas (x-axes of the graphs in section 4.5.4).

4.5.3 Recalculated individual's self-consistency and interpretations using timelines with merged units

The following section presents the self-consistency results for the MIDI PS after unit-merging. This is followed by comparisons of self-consistency results between 1) MIDI and session I performance responses, and 2) the two performance sessions (sessions I and II). This is done to a) assess the degree of self-consistency of individuals, b) assess the 'difficulty' of the pieces and renditions, and c) identify the responses that are self-consistent enough to be used for the comparison between individuals (section 4.5.4) leading to the comparison of the responses with the musical features (chapters 10-12).

4.5.3.1. MIDI, PS results

The second of graphs 4.5.3.1, appendix 4 (summarising the results in table 4.4.2.3, appendix 4) presents the proportion of self-consistency figures that show complete self-consistency, self-consistency, or fall below the self-consistency threshold when units are merged according to confusion points. The majority of responses are now self-consistent. Comparison of pre-merged and merged data (graph 4.4.2.3.1, appendix 4) shows the differences in proportion of self-consistency. The improvement of self-consistency shows that many of the 'inconsistent' cases were inconsistent because of presses that were in adjacent units (in the same area).

The results indicate the difference in difficulty in identifying positions and areas between the pieces; the more difficult the identification of areas, the lower the self-consistency (graphs 4.5.3.1). The differences in proportion of completely self-consistent responses indicate the differences in ease or clarity of the phrase structure for the listeners for each piece. The Bach Suite and Mozart Sonata seem clearest.

4.5.3.2. Comparing MIDI and performances PS (I)

Self-consistencies are also used for the comparison between responses to MIDI and performances (graphs 4.5.3.2.1-2, section 4.5.3.2, appendix 4) and between different performances in the different sessions (graphs 4.5.3.3.1-2, section 4.5.3.3, appendix 4). Graphs 4.5.3.2.1-2 show the proportion of self-consistency in the three different categories. They show that the self-consistency figures usually are higher for the performance than the MIDI responses. This may be reflecting the greater preparation/expectation of PS in the performances. There are two exceptions for the consistency results of the original time line (unmerged): the Uchida and Barenboim recordings. The situation reverses when the merged timelines are used indicating that the lower values result from the larger response areas rather than completely different and distant responses, and therefore it seems that these performances prepare the listeners to a lesser extent than the MIDI or confuse the musical cues.

There are also some differences among the consistency figures between performances. These are, for the most part, concentrated in the difference between the proportion of complete agreement, and agreement of $1.0>k\geq 0.7$. Generally, in pre- and post- merging, the consistency figures stay above 0.7 for at least 75% of listeners. The only exceptions are the pre-merged Bach Passion, Uchida (which improve post-merging implying that the inconsistency was around the unit-boundaries), and pre- and merged Mozart Aria (which has only 2 listenings), which have less than 75% above the 0.7 threshold. The responses that fall below the threshold in pre-merged responses for the performances tend to be those of N or sometimes M, but not DL.

4.5.3.3 Comparing the two performance sessions

Graphs 4.5.3.3.1-2, appendix 4 show the proportion of responses for the two performance sessions in the same three classes: completely self-consistent, self-consistent, and below the threshold of consistency. The graphs show that if the responses in the groups $k=1$ and $1>k\geq 0.7$ are taken together, the degree of consistency is more or less the same between the two performance sessions. The proportions within the completely self-consistent group ($k=1$) differ between sessions. However, the patterns are different among the pieces:

1. The proportion of complete agreement is higher for both performances in session II for the Bach Passion, Mozart Sonata and Mozart Aria.
2. Bach Suite: those who heard Gendron are high in session I and lower in II. For both, the proportion of complete agreement is higher after merging. Those who heard Rostropovich first are lower but improve after merging and in session II.

3. Brahms: The proportion of complete agreement is higher in session II for Gould but not for Kovacevich.
4. Wagner: there is a slight decrease in the proportion of those of complete agreement for both Barenboim and De Waart.

The results of previous analysis (chapter 3) indicated that the phrase-structure identified in session I sometimes affected the phrase structure identified for the second recording (session II). This is also shown here (see also chapter 11.4.3).

4.5.4 Inter-listener consistency of 'interpretations' and their groups

Inter-listener consistency is high enough to allow groups of interpretations to be built. The results are shown in two sections: 1) the three phrase parts: PS, PE and EOP for MIDI and 2) the PS for the different renditions of the same piece. For the following discussion, the table of agreement statistics, graphs and more detailed discussions are given in 4.5.4.1 appendix 4.5. With the exception of the Brahms PE's only interpretations that have agreement of at least 0.67 with at least one other interpretation are included in the groups and few interpretations are excluded.

4.5.4.1. MIDI PS, PE and EOP

Bach Passion For the PS, the responses indicate that most groups seem to respond most often to features that appear more frequently and have an idea of a phrase that is relatively short while those in one group respond to features that occur less frequently, and have an idea of a phrase that is relatively long. The common positions for all the groups are 1-1.625, 5.375-6.125, 8.875-9.625. There are fewer groups for the PEs than PSs. Areas 5.375-6, 8.75-9.5, 10.375-11.125, and 13.375-13.875, are represented in all PE groups. Areas 2.375-2.875 and 4-4.5 are represented in two groups but only chosen by a small proportion of one. The areas of PE often precede and overlap with PS areas and the majority interpretations seem to be clearer for PS than PE. Overall there are fewer positions than in the PS. For the EOPs the following positions are chosen 3.5-4.375, 4.875-5.75, 6.875-8.875, 9.9375-10.875 and 12.5-13.875. The EOP areas chosen begin before, are longer than, and overlap those of PSs and PEs. In general, the graphs indicate that the PE is the 'clearest' phrase part identified.

Bach Suite The common PS positions for all groups are 1-1.375 and 2.875-4.625. The important PE areas are 1-1.375, 2.875-3.875, 5-5.875 and 6.875-7.25, with 1.875-2.375 being chosen less often. There is more merging in the PS than PE; the PS presses occur nearer note boundaries than the PE key liftings thus reducing merging in the latter. For the EOP, the areas merged begin before, are longer than, but overlap with those of PS and PE. Overall, there seems to be less difference between the clarity of identification of the different phrase parts than in the Bach Passion.

Mozart Sonata There are two main PS groups and they have the first three positions in common (1, 2.5, 4.333). Group 2 also has 6.333. Both have complete agreement. Two PE groups identify 2.333-2.833, 4.166-4.833, and 8-9 and one also identifies 6.333-6.666. These areas begin earlier than, but coincide with the PS areas. Both groups have very high levels of agreement. The EOPs include areas 1.333-2.666, 3.333-4.333, or 4.333-5, 5.666-6.666 and 7-7.666, or 7.666-9. The interpretations of PS, PE and EOP show that they are the clearest among the case-study pieces. Area 6.333-6.666, the elided phrase however, has less general agreement (chapter 10).
Mozart Aria The interpretations of PS, PE and EOP show that there are two phrase structures, one of four phrases and the one (present in the interpretations of a third of the listeners) that subdivides the first and third phrases. Again, there does not seem to be a large difference in 'clarity' between the different phrase parts.

Brahms The phrase starts include: 1-1.666,4.666-5.333, 6.666-8.333, 8.666-9.666 (and 2.666-3.333, 12.666, 10.333 and 13.666). For the phrase ends overall the pattern is less clear than that of PSs. The majority of the listeners press in area 2.666-3.333, 6.333-7.333, 9.333-10.333, and 12.333-13.333. The difference between the positions in the interpretations of PSs and PEs in this piece indicates that there is a difference in perceived clarity between them. For the EOPs, areas 3.333-4.666, 6-7, 8.666-9.333, and 12-12.666 are the most commonly represented. Overall for the Brahms, the PS responses are the clearest and the PE and EOP are more blurred.

Wagner The majority of the PS interpretations include 1, 10, 20, 26.25, 35. Most of the PE interpretations include 8.75-10, 14.5-15.75, 20-22.25, 26.25-28, 30.75-32, 36-37.25, and 42. Unlike most of the other pieces, most of these overlap exactly with those of PS and do not precede them. Area 4.5-5.25 occurs in individual interpretations but these do not form enough of a majority to appear in a group. This is not the case for the EOP which includes areas 4-5, 7-9, 13-15.125 and 25-27.25 which are represented in the majority of groups and areas 19-20, 29-30.75, 34-35, and 41-42, which occur only in a minority. The interpretations of the EOPs show that there is a considerable difference in the level of expectation among phrase boundaries.

Summary

This analysis determined the level of listener's self-consistency in consecutive listenings and population agreement between 'interpretations'. This led to a clearer representation of the results in the form of discrete groups of interpretations of PS, PE and EOP areas and a quantification of the levels of agreement.

The results show that in each piece some areas are unanimously chosen while others are identified by individual groups of interpretations, reflecting their relative clarity. In general, the EOP, PE and PS areas overlap, with the beginning of the EOP areas slightly preceding the PE areas, and the PE areas slightly preceding PS areas. The degree of overlap and size of area of EOP, PE and PS vary within and between pieces.

4.5.4.2 Comparisons of individual raw data responses with interpretations-multiple interpretations and 'mistakes'

Looking back at the raw data, in many cases the responses to the individual listenings that were 'different' from the interpretation assigned to that listener are in agreement (to different extents) with other interpretation groups identified. Graph 4.5.4.2.1, appendix 4 shows an example from the Mozart Sonata, in which a DL, whose interpretation is the same as the majority of those of Group 2 (graph 4.5.4.1.3.1, appendix 4), has one listening in which the positions in his response are identical to the those of Group 1.

There are some positions (such as 3.5 in graph 4.5.4.2.2, appendix 4) for which far fewer listeners press keys and only once by each listener. Graph 4.5.4.2.2, appendix 4 shows the three listening responses and the interpretation for a DL for the Mozart Sonata. The interpretation assigned to this listener's responses is in Group 3 (graph 4.5.4.1.3.1, appendix 4). The listener pressed once (in the first listening) on 3.5. This is the only raw response in the whole data set at this position (i.e. between 3.5 and 3.666, chapter 3). This DL also chooses 3.833 in all three listenings. This comparison indicates that the 3.5 is an anticipation of 3.833 which is chosen for all three listenings. A comparison with the rest of the raw data shows that this is the only listener who chooses 3.833; this listener is self-consistent here but does not share it with others.

These examples illustrate: 1) the ways in which this method may be used in order to distinguish between 'intentional' and 'mistaken' responses in such a free response method and 2) that though interpretations may be identified, the same listeners' responses sometimes fit with more than one interpretation group – indicating listeners' variety of interpretations.

4.5.4.3 Comparison of MIDI and performances PS

Responses to the different rendition types and two performance sessions were also compared. The comparison shows that in general, the areas identified by the different groups are very similar though there are some differences.

In the **Bach Passion** there is more merging in the MIDI than in the performances. The solo line and accompaniment are sometimes not together and the higher level of merging in the MIDI may be because here this causes confusion whilst in the performances the solo line is highlighted.

In session I of the **Bach Suite** the interpretations in response to Gendron have fewer positions that the Rostropovich and the MIDI responses are in between, indicating that Rostropovich highlights more phrase areas than Gendron and that the same boundaries are recognised in the MIDI rendition. The interpretations of session II seem to be affected by the first (chapter 11.4.3). This relates to the higher self-consistency figures of the Gendron (section 4.5.3.3)

In the **Mozart Sonata,** the PS interpretations for the MIDI and performances are most consistent both within and between performances among all the pieces. All include the areas 2.5-3 and 4.333-5. Some of the MIDI interpretations also include 6.333-6.833 as do almost all of the Uchida responses in session I but only about half of the interpretations for session I of the Lipatti include this position. As in the other pieces, there seems to be an effect of session I on the second; in session II of the Uchida only about half of the interpretations include 6.333 to 6.833 while in the Lipatti the majority now include this position (chapter 11.4.3).

In the **Mozart Aria** and **Brahms** the proportions for the areas differ depending on the rendition and session. For both, there are two groups of areas. One includes the longer phrases and the other also includes the shorter ones and, as in the previous pieces, there seems to be an effect of session I on the responses in session II. The same trends are seen in the **Wagner**, the same positions are identified but with different proportions for different renditions, changing between sessions (4.5.4.3, appendix 4 gives the graphs and a more detailed discussion).

The comparison between MIDI and performance interpretations shows that the areas identified in response to both are the same. Comparing the responses to performances in session I shows that there often is emphasis on more positions in response to one rendition than the other. The comparison of both sessions indicates that the responses to II are influenced by I. The results also show that most ability groups are represented in most interpretation groups.

4.6 Summary of statistical study of listeners' self-consistency and population agreement

The method was adapted to the analyses of listeners' individual phrase responses in order to determine: 1) listener's self-consistency and 2) on the basis of interpretations concluded through this process, the population agreement of the identification of PS, PE and EOP for MIDI and performances. This was done because of the lack of a published standard method for that purpose. The method was found to be successful in quantifying the degree of agreement on phrase part areas and the similarity between interpretations in the different case-study pieces. The summary of the results of this section overlaps in part with, and strengthens, conclusions drawn from, other analyses of the same data reported in chapter 3.

This method helped to identify that:

1. Most of the listeners were significantly self-consistent. Moreover, differences between the listenings may be a result of deliberate choices. This becomes even

clearer when comparing the raw responses to the grouped 'interpretations': though a listeners' overall 'interpretation' can be categorised in one group, responses to single listenings can fit with others. Positions may be combined in several different ways, even by the same listener, indicating primacy of positions over position combinations.

2. There were two sources of inconsistency in the raw responses: the listeners pressed in completely different positions during different listenings or positions within an area.

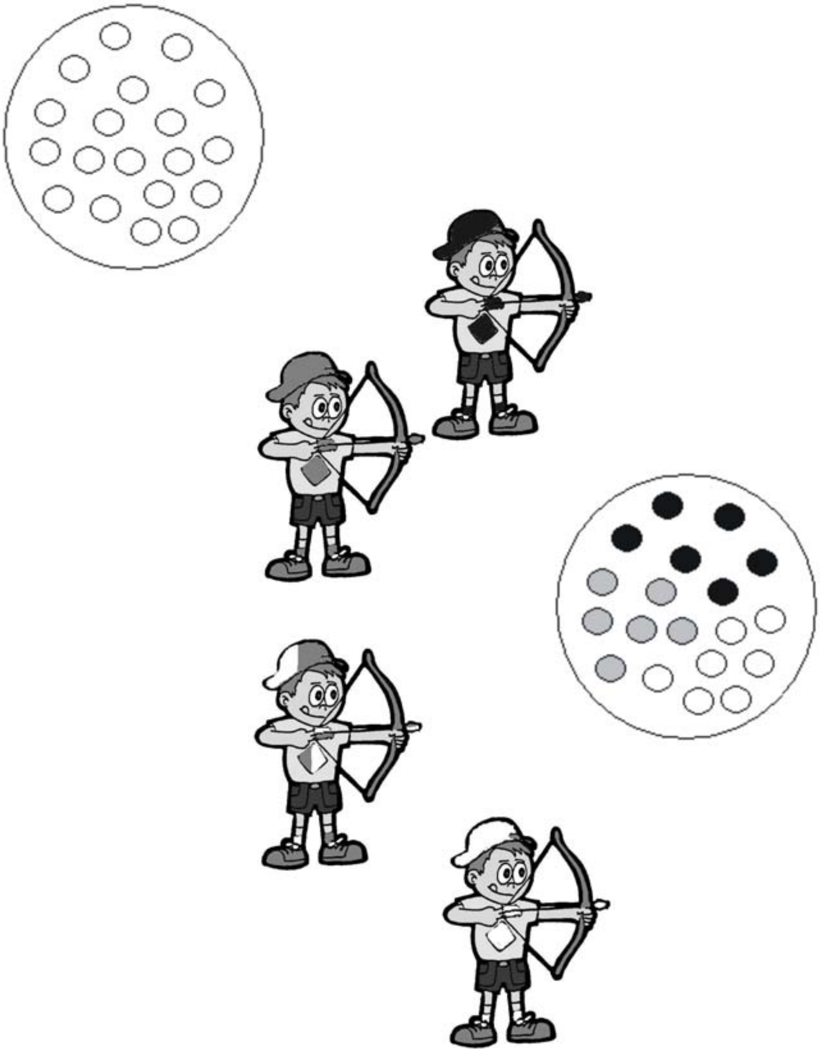
3. A method was developed in order to distinguish between the two possibilities by merging units that included close-to-note-boundary key-presses. This allowed the systematic identification of areas of PS, PE and EOP.

4. The merged data was used to quantify listeners' self-consistency (which was found to be significantly self-consistent in most cases) to represent individual listener's responses as a single interpretation and to compare between these interpretations.

5. The comparison between interpretations of the same rendition showed that some interpretations representing different listeners' responses showed significant agreement. Several interpretations could be assigned to a single group (not dictated by amount of musical experience). Furthermore, this allowed for different areas in the piece to be identified as 'significant' PS, PE and EOP. This allows for a comparison between different groups of interpretations, areas identified as PS, PE and EOP, and renditions.

6. The comparison between the different renditions showed that, in general, the same positions are chosen in response to the different renditions. The interpretations of the MIDI renditions usually include all the areas identified in the performances. There are rare occasions of interpretations that occur in the performances but not in the MIDI.

7. The proportions of responses often change depending on the rendition and session (I or II). The proportions of MIDI responses are often intermediate, with the proportion of responses for the different performances being higher and lower than the MIDI. Moreover, in session II, there are often effects of I. Among the interpretations of performances two types of differences were recognised: a) Two related, but clearly different interpretations such as in the Mozart Sonata, Aria and Brahms. These share common areas while some interpretations have additional ones. b) A number of common areas but also different interpretations have combinations of different additional areas such as in the Wagner. These results form a basis for discussion in chapter 10. Before this, having explored the boundary areas and the way in which they relate to one another, in the following chapter a more local view is taken concentrating on the nature of the boundary areas.



The relation between phrase part interpretations, their groups, and the affiliation of their identifiers. Note the possible multiple affiliations.

Chapter 5

Clicks and Phrases - Reaction and Recollection: Boundaries

5.1 Introduction

Much of the current study has so far approached the question of phrase identification from the large scale to details; the observation of listener responses has concentrated on broad areas of the music. In this section the opposite approach is taken; responses to short excerpts from three of the case-study pieces are studied in order to explore the nature of the phrase ends (PEs) and phrase starts (PSs) in more detail; primarily the "exact" location of phrase "boundaries".

This study uses the responses to clicks, short 'characterless' sounds equivalent, in their importance relative to the rest of the stimulus, to noise. These are superimposed at different positions around phrase boundary areas identified in the previous studies (chapters 3 and 4) and suggested by music analysis (chapter 10). This is a very small study, with only three examples and a small number of listeners. It produced some promising findings that could lead to further work.

Previous experiments that use a click superimposed on linguistic or musical stimuli next to phrase boundaries have done so through two approaches:
1) Reaction - listeners are asked to react to a click immediately on hearing it (studies include, Abrams and Bever, 1969; Bond, 1972; Flores d'Arcais, 1978; Holmes and Forster, 1970), and

2) Recollection - listeners are asked to remember where the click occurred and then to indicate its position during second hearing (or reading) without the click (for example, Stoffer, 1985).

These studies have found that for 1) Reaction - the reaction time is shorter for clicks placed at the phrase boundary than further away at either side of it and for 2) Recollection - the recalled click position tends to 'migrate' to the phrase boundary. This phenomenon is often referred to as 'click migration'. In some studies the two approaches are combined; listeners are asked to mark the location of the click either immediately or at the end of the segment on a score or visual representation (Kaminska and Mayer, 1993).

These observations have been related to a theory that we respond to extraneous noise more quickly when the "cognitive load" of what we are concentrating on is smallest and that we "wait" for a relatively low cognitive load to "deal with" the additional information (hence the click migration). These theoretical and experimental approaches were developed in studies of processing of music and language, primarily for the exploration of responses to segment, mainly syntactic, boundaries (for example, Fodor and Bever, 1965; Kaminska and Mayer, 1993).

The conclusions of previous studies are taken as a basis for this study and it is therefore expected that perceived phrase boundaries can be identified using this approach. A detailed explanation of background and methodological aspects is needed to set the basis for the experiments and the interpretation of the results. Some previous experiments and conclusions from both language and music studies will be discussed here followed by a summary of some of the methodological aspects.

Aims

1) To evaluate the applicability of this approach to the study of phrase boundary perception in excerpts from the western classical repertoire.
2) To obtain an indication of the perceived location of phrase boundaries.

5.2 Background

5.2.1 Unit integrity, information processing, cognitive load and click detection

Click studies are based on a number of assumptions. One is that there is a tendency in perception to preserve the integrity of a perceptual unit (in this case the phrase) by resisting interruptions. The click is used as an interrupting stimulus which the listener is required to locate relative to the perceptual object (in this case the perceived phrase "boundary"). The expected clustering of the clicks towards the boundary in perception is taken as an indication of its perceived location and the strength of this perception.

It is assumed that 1) the higher the cognitive load, the longer the reaction time to the click, and 2) that clicks 'migrate' in our memory towards the end of segments of information. In terms of phrasing, the cognitive load is considered higher during the phrase than at the phrase boundary (for example, Gregory, 1978). Clicks superimposed at the phrase boundary should be responded to most quickly (reaction) and their location should be most accurately remembered (recollection). Clicks further away from the phrase boundary should be responded to more slowly (reaction) and should migrate in the listeners' memory to the nearest phrase boundary (recollection).

Listeners are assumed to attempt to incorporate new events into their representation of the piece. In some cases the new event (in this case the click) is incompatible with the mental musical grammar. If a section of the piece has just finished, then this incompatibility has little effect, reaction to the new event is quick, and its location correctly remembered. If the new event arrives in the middle of the section, it cannot be dealt with quickly, the reaction is slow and its location is not easily remembered. Furthermore, in the latter case the theory is that we "wait" until the end of the section to deal with the new event and therefore, in retrospect, we remember having heard it at the end of the section. The range of grammatical sections for which this phenomenon occurs seems to be large, including clauses and sentences in language, and bars, prolongation structures, modulations and phrases in music.

Since click identification is a relatively simple task, it is likely to cause the listener only a minor distraction from listening to the music and therefore should provide a sensitive reflection of the cognitive load resulting from the musical processing (Berent and Perfetti, 1993, p. 207). Click detection has been found to be sufficiently attention demanding to create interference with a variety of primary tasks such as matching letter transformations and lexical decisions in judging homophones (Kellas et al., 1988; Posner and Boise, 1971; Posner and Klein, 1973).

There are several aspects of this approach that make it attractive for the current study. As in the other listeners' study (chapter 3), they allow the exploration of real-time parsing decisions while not terminating musical processing responses and no verbal or visual communication between the subject and the experimenter are necessary. Moreover, listeners are not required to actively identify PE/PSs. On the contrary, it should enable the inadvertent 'identification' of the exact location phrase "boundaries". As will be shown, it does not require musical experience.

5.2.2 Click studies and music

These have been used in the investigation of a number of different aspects of music in which cognitive load should differ under different circumstances (such as comparison of cognitive load between unprepared chromatic modulations and non-modulating passages, Berent and Perfetti, 1993).

Other click studies (reaction and recollection) have investigated boundary identification in music. For example, Kaminska and Meyer investigate click migration to metrical and intonation boundaries (1993, p. 155). They used melodic lines, composed of isochronous notes and conforming to the same basic structure; a metrical boundary dividing two melodically-identical but pitch different phrases. Musically untrained subjects were asked to indicate the location of the click using a schematic visual representation of the tune, either immediately, as soon as they thought they heard the click, or retrospectively, at the end of the line.

The clicks tended to be localised closer to a boundary than their actual position in the melody, implicating both grammar and intonation, acting independently or in combination in perception (Kaminska and Mayer, 1993, p. 157). The '[i]nformation about phrase is embedded in the total metrical structure. The whole has to be appreciated before a decision as to how to parse it can be made; early decisions would be too error-prone. Intontational information, on the other hand, is carried at the surface level, is intrinsic to the ongoing acoustic input, and is available immediately as a travelling wave as the sequence unfolds' (1993, p. 157) in manner similar to that in speech perception (1993, p. 157-8). '[L]istening to music is by no means a linear, data-driven unfurling of auditory events.… The parallels of divergence between stimulus parameters and cognitive representations in speech and music signal at the theoretical level, and substantiate at the empirical, the constructive processes involved in listening to music' (1993, p. 160). This questions 'the traditional division made on the basis of differential relative weighting of bottom-up to top-down processes' (1993, p. 160). '[T]he psychological world of music is not necessarily in complete harmony with the physical world. There exists a considerable degree of freedom in the conscious realisation of music, and what is heard may not be so much an echo of the physical dimensions of sounds as a subjectively generated variation on the theme' (1993, p. 160).

More recent click studies by Martinez investigated the 'prolongational structure of tonal melodies' (Martinez, 2002, p. 633).[36] The hypothesis was that 'clicks located at the prolongational boundary will not migrate while clicks located before and after the prolongational boundary will migrate to the boundary' (2002, p. 633). The melodies were from western tonal art music and professional musicians listened to each melody three times first without then with the click, and then without the click again, this time pressing a key when they believed the click had occurred (2002, p. 633). Differences between responses to click at different

[36] 'Prolongation is a structural phenomenon described in music theory, in which some pitch events remain active within the musical 'flow' even though they are not physically present' (Martinez, 2002, p. 633).

locations were not found to be significant, though those of a previous study were (Martinez, 2001a; Martinez, 2001b both reported in; Martinez, 2002, p. 633).

5.2.3 Alternative methods

Click detection (both reaction and recollection) is just one of several techniques that could be used for the more precise location of perceived PE/PSs. Others include the 'probe tone' technique developed by Krumhansl and Shepard (1979) and employed by Krumhansl and Kessler (1982) in a study of the dynamic changes in the representation of a modulating sequence of chords. A variant for phrase boundary perception could be to stop the music at different positions and ask for a rating or yes / no response as to whether the phrase had finished or not. Although methods based on explicit probing or stopping the music and questioning can provide valuable information regarding the representation that the listener has constructed at each of the times of probing or questioning, such techniques have some limitations including that the demand to provide an explicit judgement necessarily terminates the listening process and might encourage representation commitments that might have otherwise been suspended temporarily (Berent and Perfetti, 1993, p. 206).

5.2.4. Click detection and phrasing

A small number of studies have explicitly used these methods for the investigation of 'phrasing'. Gregory found that there was a significant tendency for the click to be attracted to phrase boundaries (Gregory, 1978, p. 171).[37] However, his definition of 'phrase' relies on the different ways a series of six notes are stemmed and beamed and thereby grouped visually – in twos or in threes. Sloboda and Gregory, referring to this work, do not use the term phrase; 'Gregory (1978) has demonstrated that a click presented during a six note musical fragment tends to be perceived later than its actual time of occurrence, and that the perceived temporal location depends partly upon the way the fragment is notated for the subject. The click migrates perceptually towards a boundary between two note groups' indicated by the beaming (Sloboda and Gregory, 1980, p. 274).

However, Sloboda and Gregory (1980) identify a number of drawbacks with Gregory's experiment: 1. 'the fragment made equal (and rather little) musical sense whether conceived of as two groups of three or as three groups of two. Although this experimental imposition of segmentation produced click migration, it does not necessarily follow that listeners will spontaneously segment 'real' music when not supplied with such explicit segmentation cues.' 2. 'the segmentation was unrelated to any rule system that might be held to govern the construction of a melody (1980, p. 274). Stoffer (1985) has similar objections: 1. Visual segmentation on the score might have been the only cue for cognitive

[37] Although this effect is significant, it is not very marked. This may be partly because not all of the subjects perceived the phrases in the way suggested (Gregory, 1978, p. 173).

segmentation. 2. The sequences did not exhibit musical regularity that could have been recognised by a listener. Without that, no top-down controlled cognitive segmentation is possible that will go beyond the mere detection of chunk boundaries formed by a bottom-up analysis of the sequence. 3. The click localisation method is susceptible to response biases. A systematic displacement that is dependent on syntactic structure and independent of response biases occurs only when subjects are primarily attentive to the stimulus pattern in one ear and not to the click in the other. There does not seem to be agreement about whether listeners actually perceive the clicks at the displaced positions. They may not perceive any displacements, but rather produce some kind of a response bias that does not concern the form of the task as discussed here (Stoffer, 1985).

If the click location is the primary task, as was the case in Gregory (1978), it may be that the listener waits for the click, and only then attention was switched to the musical structure. In that case, click localisation would always be late (Stoffer, 1985, pp. 194-5). Therefore, location of the click should be a secondary task.

If these objections are correct, Gregory's (1978) results show only an effect of visual grouping of the notes on click localisation (Stoffer, 1985, p. 195). The first and last objections also apply to the click localisation experiment reported by Sloboda and Gregory (1980) (Stoffer, 1985, p. 195).

In order for a click localisation experiment to be immune from these objections:
1) Subjects should not be allowed to read the score until the click is detected.
2) Attention must be focused primarily on the musical structure by asking subjects to perform a task that forces them to attend to the music.
3) The musical material should exhibit genuine musical regularities that can then function as phrase markers e.g. change in melodic contour, melodic regularities formally described as transformations, harmonic progressions, especially cadences, rhythmic regularities, patterns of pauses, and relative note durations (Stoffer, 1985, p. 195).

Stoffer (1985, Experiment 2) explored the effects of different 'phrase' structures on click detection. This study examined the adherence to a binary heuristic in segmenting melodies as a function of listeners' musical expertise and experimental training. Subjects at two levels of musical expertise who had been trained to discriminate binary from ternary phrase structures were presented with melodies of these two types and their results indicated that in both binary and ternary structures, responses to clicks occurring on a first-order boundary in the first half of the melody were faster than responses to clicks at any other position. In the first half of the ternary structure responses to clicks located at the second-order boundary were faster than those at the third-order boundary. Stoffer's results demonstrate that the click detection task provides a reflection of listeners' representation of the hierarchical phrase structure of a musical piece.

Returning to Sloboda and Gregory (1980), for them, phrase boundaries are denoted by structural and/or physical markers. They take the musical phrase as a given and investigate whether it can be shown to be psychologically real. When talking of 'phrases' or 'phrase boundaries' they mean units that trained musicians identify by consensus, not structures that are completely defined, either formally or psychologically (Sloboda and Gregory, 1980, p. 275). Their method combines reaction and recollection; using a visual response method, subjects could mark their responses on the score anytime during or after listening (Sloboda and Gregory, 1980, p. 276).

Sloboda and Gregory found that clicks tended to migrate towards phrase boundaries, paralleling results obtained with language. When phrases were marked physically (retaining contour but destroying harmonic sense), clicks occurring both before and after a boundary migrated towards that boundary. When phrases were marked structurally (preserved harmonic sense and longer notes at the end of phrases), only clicks occurring after a boundary migrated towards it. They conclude that both physical and structural phrase markers affect migration. That structural markers seem to exert an influence on click location only if the click comes after the phrase boundary suggests that subjects were unable to anticipate the phrase boundary on the basis of structural cues alone. This may have been because they did not read through the melodies before hearing them. In contrast, the physical marker, a longer note, seemed to elicit anticipatory migration. This could be because it was highly salient from a cursory visual inspection of the melody and could be used as an 'anchor point'. There was a third condition where no markers were present, which showed a small migration effect.[38] This may be because the melody retained the rhythmic structure of the marked conditions, and this may have given some residual cue for grouping (Sloboda and Gregory, 1980).

In contrast to language studies, there was an overall tendency for clicks to be perceived later than their actual time of occurrence (Sloboda and Gregory, 1980, p. 274) an observation that has been made in several experiments (including Fodor and Bever, 1965; Gregory, 1978). One explanation is Titchener's (1909) law of prior entry: subjects are attending to the melody and so it gets processed first (Gregory, 1978). This account leaves the tendency for early perception of clicks in speech unexplained. An alternative explanation is that subjects do not perceive a click occurring in the middle of a note as synchronous with it. For two percussive sounds to appear simultaneous, their onset times rather than centres are expected to be synchronised. It is likely that the subjects perceived onset asynchrony as evidence that the click came after the note, even though it occurred while the note was still sounding. Sloboda and Gregory's results were recalibrated to take account of 50 ms asynchrony, and found that subjects perceived the click as occurring on average 14 ms earlier than it actually did. This account reconciles the differences

[38] This condition was the same as the 'physical marker only' one except that the crotchet at the phrase end was replaced by two quavers (Sloboda and Gregory, 1980, p. 276).

between results of language and the music studies (Sloboda and Gregory, 1980, p. 279).

5.2.5. Language and click migration

Although click recollection was first proposed by Ladefoged and Broadbent (1960) with the aim of studying perception of temporal sequences, it was accorded most attention in speech perception in the late 1960s. Fodor, Bever and Garrett (in studies such as Fodor and Bever, 1965; Fodor et al., 1974; Garrett et al., 1966) showed that clicks were subjectively attracted toward clause boundaries and that their location was reported more accurately when clicks coincided with major syntactic breaks. They suggested that during sentence perception, clauses function as perceptual units resistant to click intrusion (Fodor et al., 1974). '[W]hen a click is sounded during auditory presentation of a sentence, a subject is most likely to report its location correctly if it occurs in the major grammatical (i.e. clausal) break, and that when it comes at some other point in the sentence, erroneous judgements of its true location tend towards placing it in the grammatical break, or else in positions adjacent to it' (Fodor et al., 1974).

Fodor and Bever conclude that: i) Clicks are attracted towards the nearest major syntactic boundaries. ii) The number of correct responses is significantly higher in the case of segments. iii) These results are consistent with the view that the segments marked by formal constituent structure analysis function as perceptual units and that the click displacement is an effect which ensures the integrity of these units. iv) The distribution of acoustic pauses in the sentential material does not account for the observed distribution of errors. v) There is a slight tendency to prepose responses to clicks in sentences. This tendency is reversed during later stages of the experimental session. Both of these effects are asymmetrical for the two ears (Fodor and Bever, 1965, p. 414).

Many of the musical studies compare their stimuli and results to structures in language and the results of language studies. For example, Kaminska and Mayer take the musical counterpart of grammar in speech to be metrical structure and that of spoken intonation to be performance intonation (Kaminska and Mayer, 1993).

5.2.6. Language and click detection

In the late 1970s psycholinguists became interested in on-line procedures that could help uncover ongoing sentence comprehension processes. The click-monitoring method (here referred to as recollection) was then criticised and abandoned because subjects answered long after having heard the sentences and therefore their responses might not be faithful reflections of perceptual processing. In an experiment in which subjects were encouraged to respond even when no click was actually present, Reber (1973) showed that subjects tended to localise nonexistent clicks at syntactic boundaries.

Instead the methods referred to here as 'reaction' were used. Studies that used this procedure (Abrams and Bever, 1969; Bond, 1972; Flores d'Arcais, 1978; Holmes and Forster, 1970) suggest that reaction times are shorter in the first than in the second part of sentences and to clicks at major syntactic boundaries than at minor breaks or within syntactic constituents (discussed in Cutler and Norris, 1979 and Cohen and Mehler, 1996).

5.3 Methodological aspects

As indicated by the above studies, there are many variants of this method. Here, some specific methodological aspects of the current study are discussed with reference to the literature discussed above.

5.3.1 Reaction and Recollection

The experiments investigating 'phrasing' have involved a combination of reaction and recollection tasks, with most studies asking listeners to mark the location of the click during or after hearing the extract. In this study, both reaction and recollection methods are used but are clearly separated. Furthermore, in most previous studies click location was marked on a score or another visual presentation. Here, all responses are within the auditory domain.

5.3.2 Length and type of extract

The extracts used for these studies are short (and sometimes do not reach lengths often associated with "phrases" found in other studies and in the current one). Kaminska and Mayer, investigating click migration to metrical and intonation boundaries use 'phrases' of approximately 5 seconds duration (1993) and those used by Gregory are 6 quavers long (1978). Most of the examples, and all those studying "phrasing" were composed specifically for the experiments and very simple structures. Only Martinez (2002) uses examples from the western classical music repertoire.

In this study, three of the case-study pieces (see chapters 3 and 10 and sections 5.5 and 5.6) are used to obtain an indication of the location of perceived phrase boundaries, to evaluate the precision of the results in comparison with those obtained in the listeners' phrasing study, and to test the relationship between the responses obtained in this study and the phrase-type categories identified in the excerpts.

5.3.3 Click characteristics and positioning

To avoid masking of the click by the note or vice-versa the two have to be played one after the other. However, if two note-onsets are too far apart, they are not considered to have occurred together by listeners and therefore the question of may become unclear (Sloboda and Gregory, 1980, see section 5.2.5 above). Kaminska and Mayer use a click of equal volume to the notes superimposed in a pre- or post-boundary position, the click-to-boundary distance being constant (Kaminska and Mayer, 1993). More specifically, Berent and Perfetti place the click 100ms after their triad. There is an interval of at least 450 ms between the onset of the triad preceding the click and the onset of the next musical event (Berent and Perfetti, 1993, pp. 212-3). In this study, the click was placed within this time window and the listeners were asked to identify the note *during* which the click was heard.

In this study, the position of the click was varied for each listening. All were within one beat and a quaver before or after the boundary area. Depending on the number of intervening notes there were between five and seven click positions. The listeners heard the clicks at one of these positions each time.

5.3.4 Distracter tasks

The click tasks should be combined with another (distracter) task so that listeners are forced to listen to the whole extract and not concentrate purely on the mechanical position of the click (Stoffer, 1985, section 5.2.5 above). Berent and Perfetti use a melody recognition task; listeners were given a memory probe after they had been presented with a number of extracts and asked to identify whether or not it had been heard in the previous block (1993). Kaminska and Mayer ask for ratings of pleasantness and musicality of the melody (1993, pp. 155-6). In this study, listeners were asked one of three questions (see below, section 5.4.4)

5.3.5 Ear of presentation to the listener

The music and click may be presented together in both ears, or the click in one and music in the other. Studies have shown that the positions of the clicks are judged differently if they are presented to the left or right ears (Gregory, 1978). These results are similar to those from click experiments for speech, though the late judgments in music contrast with early judgements in speech (Gregory, 1978, p. 171). However, according to Sloboda and Gregory presenting the stimuli to one ear or the other made no significant difference to the response (1980, p. 277). In this study, both the click and the music were presented in both ears.

5.3.6 Format of presentation to the listener

Both the reaction and recollection tasks may be carried out with and without scores or another visual representation. For example, Sloboda and Gregory ask their subjects to mark their responses on the score anytime during or after listening (1980). Kaminska and Mayer asked their subjects to indicate the location of the click using a schematic visual representation of the tune, either immediately, as soon as they thought they heard the click, or retrospectively, at the end of the line (1993). Gregory's study, which used two different visual representations of the same notes showed the importance and influence of the layout of the score (1978).

It is unclear what advantage can be gained by presenting a score or alternative visual representation to the listeners. At the same time, the score can introduce a number of factors unwanted in this study. For example, the visual representation may itself introduce a particular segmentation (Gregory, 1978). Therefore, in the present study listeners were not given any visual representation.[39] Instead, listeners were played the same excerpt twice, once with the click superimposed and the second time without. They were asked to press a key at the moment they heard the click during the first listening (reaction time), and, during the second listening, to press a key again at the position where they thought they heard the click the first time (recollection), thus allowing all responses to remain in the musical auditory domain. This variant method does not seem to have been used in any of the phrasing studies cited above though Martinez used a similar approach in her study of musical prolongation (2002).

5.3.7 Musical experience of the listeners

Subjects from a number of different musical backgrounds have participated in the studies discussed above (section 5.2). In several only one group has been studied. For example, Kaminska and Mayer studied only musically untrained subjects (1993) and Martinez uses only professional musicians (2002). Stoffer studied subjects with two levels of musical expertise. However, they had all been trained previously to discriminate binary from ternary phrase structures (1985).

The listeners in this study were the same as those of the studies discussed in chapter 3: Degree Level Musicians (DL), Musicians (M) and Non-Musicians (N). If there is a difference between the groups, the distance of reaction and migration are expected to be more pronounced in the Ns as they may not have developed the strategies and practice of locating sounds in a musical stream. DLs are expected to have already acquired strategies for remembering specific notes in a stream.

5.3.8 Analysing the responses

Some studies have automatic limits to identify 'hits', 'misses' and 'false alarms'. For example, for Berent and Perfetti (1993), the lower boundary of a hit response was taken as 100ms, the assumption being that shorter response latencies resulted from errors of anticipation. The upper limit of the hits category was based on the reaction time distribution by cutting the distribution at the point where it became

[39] At the end of some of the listenings, listeners were presented with a score after hearing an example with a click and asked to mark the score. This was only done with the musicians and, because of time restrictions, a small number of listeners were included. In general, they either marked the position at which the click was played or at the phrase end or start. A larger sample is necessary for further analysis and conclusions.

flat – 900ms. Hence, hits were defined as the first response given using a legal key within the boundaries of 100-900 ms. Misses were of three kinds: an absence of a response, a slow response (a first response with a reaction time above 900 ms), and a first response collected using an 'illegal' key. False alarms were either fast responses (a first response whose reaction time was shorter than 100 ms) or secondary responses - responses occurring after the first one. They excluded key-presses that were outside the 100-1,000 ms range after the click (Berent and Perfetti, 1993, p. 213).

Here, the only responses that were removed were those that fell as outliers in comparison to the group as a whole. Those remaining were, for the most part within the range described in these studies. No responses were 'automatically' removed as those that fell outside these boundaries could be, and some were found to be, informative (section 5.5.2.3.2 below).

5.3.9 Click detection embedded in the other study

This study was embedded within the MIDI listeners' study (chapter 3). All the tasks for the same piece were carried out one after the other. For the three pieces that had the click tasks, each started with a click reaction and recollection task so that listeners would have no prior decisions made about the phrasing from within this study. They then heard the same extract again with the click in a different position. They then heard two more such pairs: one between the other two tasks of the session (PS/PE and EOP identification) and one at the end of the tasks for that piece. For the last two pairs, listeners would have already made decisions about the phrase boundaries.[40]

Most listeners do not seem to make the connection between the two types of tasks. Most reported that they thought the click task was a distraction and that the distracter questions mentioned above were subject of the experiment.

5.4 Method

5.4.1 The pieces

Three MIDI excerpts from the case-study pieces were used. These were shorter than those of the other phrasing tasks (chapter 3): Bach Suite, bars 1-4, Mozart Sonata, bars 1-5, and Brahms, bars 1-10.[41]

[40] For those that carried out the written task, this occurred here.
[41] The procedure is given in appendix 3.2.

5.4.2 The clicks

The click positions were spread within the beat and a quaver before and after the boundary areas suggested by music theory and the experimenter, and later supported in the results of the other listeners' and performers' studies. The positions are marked on the musical examples in the results section below (section 5.5). The clicks took the form of sine-waves of 5ms duration.[42] The click positions were present in three different orders to three different subgroups of listeners to see if experience of the task or better knowledge of the piece or the other tasks had influence on the responses.[43]

5.4.3 Listeners and equipment

The experiment was run on the same equipment as for the other studies and, as mentioned above (section 5.3.7), the listeners were the same as in the MIDI study (chapter 3).

5.4.4 Instructions

Listeners were told that they would hear the same extract twice; the first time there would be a click superimposed and the second there would not be. They were asked to press a key when they heard the click during the first time they heard the extract (Reaction). During the second time they heard the extract, they were asked to press a key during the note during-which they had heard the click the first time (Recollection). Each time they were also asked one of the distracter questions:

1) How long (in seconds) is this extract?
2) Is the first note of the extract higher, lower or the same pitch as the last?
3) Is this extract longer or shorter than the extract you heard for the previous task?

5.5 Results

The results were first analysed for group differences according to order of click presentation to check for learning/fatigue effects. The results were then analysed for group differences between DLs, Ms and Ns and between those with self-reported Absolute Pitch (AP) and the rest of the groups. ANOVA were used to compare groups of responses and paired sample T-tests were used to compare means of pairs of click positions.[44]

Having seen that there were very few significant differences between groups for any of these comparisons, the responses were all analysed together by comparing

[42] Many thanks to Joel Swaine who constructed the clicks and helped with the set up of this but all of the listening studies.
[43] Many thanks to Isabel Martinez for an interesting discussion about methods of click studies.
[44] Many thanks to Dr. Vanessa Didelez for her advice about the statistical tests.

the responses to the different click positions, and then analysing the results in comparison with the musical characteristics of the extracts. For both Reaction and Recollection responses, graphs of average time and standard deviation were plotted.

5.5.1 By groups

5.5.1.1 Ordering and learning or fatigue effects

During the experiment the different pieces were presented in three different orders to the listeners constituting three groups of responses. The responses of the three different groups were compared. Box plots were plotted and outliers were removed.

One-way ANOVA showed that there was no significant difference between the groups for any of the pieces. This indicates that the order in which the click positions were heard, or the fact that some were heard late on in the experiment with plenty of opportunity to 'learn' the piece, had no effect on the responses. So, in this respect, all responses can be treated as one group. There was one exception; there was a significant difference for the reaction response position 3 ($F(2, 23) = 3.652$; $p < 0.05$).

5.5.1.2 Musical experience

The responses from listeners from different levels of musical experience were compared. Graphs 5.1.1-5.1.6 in appendix 5 show that there is little difference among the groups. This was checked statistically (after the removal of outliers).

5.5.1.2.1 Bach
Comparison between DL, M and N

One-way ANOVA was carried out and showed that there is only one click position for which there is a significant difference: reaction position 6 ($F(2, 22) = 3.122$; $p < 0.1$).[45] Responses to this position were investigated further with t-tests (see also box plots 5.1.1-2, appendix 5).

DL and M	DL and N	M and N
Difference not significant	$t(19) = 2.338$; $p < 0.05$	$t(8) = 4.266$; $p < 0.01$

These indicate that the N significantly different from the DL and M. The DL and M have a slower reaction time than N for pos 6 reaction (see appendix 5, graph 5.1.1). The average reaction time for DL is 0.43s, for M is 0.42s (for both DL + M

[45] For all these tests, only click positions with at least three responses per group were included. Here, for example, position 1 recollection (with no M and 2 N responses) was excluded.

together, average =0.42) and for N = 0.33s. There is, on average approximately a 0.1s difference in response times between DL and M, and N. It may be that at this crucial position, the 'actual' end of the phrase following the 'prolongation', the N respond to the phrase end cues with less questioning than the DLs or Ms.

Comparison between AP, [DL+M → DM] and N

Listeners with AP, and the rest of the groups were also compared. As no significant differences were found between DL and M, these were grouped together (DM). The comparison was therefore one between AP, DMs and Ns. There is only one position with significant difference between the groups, reaction position 4 ($F(2,22) = 0.169$; $p < 0.01$). The t-test shows that there is a significant difference between AP and DM $p < 0.01$. As there is again only one position with a significant difference, all responses to the Bach Suite were treated together according to this criterion.

5.5.1.2.2 Mozart

Comparison between DL, M and N

One-way ANOVA shows that there are some click positions for which there is a significant difference (see also box plots 5.1.3-4 in appendix 5):

Position 1		Position 2		Position 4
Reaction	Recollection	Reaction	Recollection	Recollection
$F(2, 25) = 7.041$; $p<0.01$	$F(2, 16) = 6.242$; $p<0.01$	$F(2, 18) = 4.246$; $p< 0.05$	$F(2, 20) = 7.932$; $p<0.01$	$F(2, 13) = 7.943$; $p<0.01$

Paired-sample t-tests were carried out to compare each pair of groups. There are no significant differences between DL and M, which suggests that they should be grouped together. The following table presents the t-test results at $p < 0.01$, $p < 0.05$, and $p < 0.1$ significance, comparing DL and N, and M and N.

Comparing DL and N				
Position 1		Position 2	Position 4	Position 5
Reaction	Recollection	Recollection	Recollection	Recollection
$t(20) = 3.103$; $p < 0.01$	$t(14) = -3.129$; $p< 0.01$	$t(17) = -1.856$; $p<0.1$	$t(10) = -4.088$; $p< 0.01$	$t(20) = -2.413$; $p< 0.05$

Comparing M and N		
Position 1		Position 4
Reaction	Recollection	Recollection
$t(10) = 3.494$ $p<0.01$	$t(7) = -2.354$ $p<0.05$	$t(6) = -2.853$ $p<0.05$

For the Recollection responses almost all the responses fall within a crotchet. Those that do not are N for position 1 and M for position 2.[46] For N of position 1, recollection, there is a range from – 0.28 to 2.46 in bar proportions around the original click position while the DL range is much smaller. This indicates that both DL and N may use the phrase boundary as their 'anchor' for the positioning of the click. The DL may be better at relating the click position to the boundary thanks to their greater musical experience. N on the other hand may remember 'on a phrase boundary' vs. 'not on the phrase boundary', with a vaguer idea of which boundary. For reaction position 1, the N are fastest than the DL and M. For recollection position 1, the N respond later than the DL and N. The same is the case for recollection position 4 and much less for positions 2 and 5 (for 2, M are slower than both).

These results indicate that, on the whole, there are not significant differences for most positions for the Reaction. For Recollection there is more difference.

Comparison between AP, DM, and N

For the reaction responses, there is only one position for which there is a significant difference between AP, DM and N: position 1, $F(2, 23) = 16.150$; $p<0.01$. The t-test shows that the main difference is between the DM and N ($t (17) = 5.463$ $p<0.01$, discussed above) and only slightly between the AP and the other groups: AP and N $t (12) = 2.790$ $p< 0.1$, AP and DM $t (17) = -2.742$ $p < 0.1$

For recollection, there are more significant differences:
AP DM N Position 1 $F(2, 19) = 7.362$; $p<0.01$,
AP and DM $t (13) = -2.904$ $p<0.1$, AP and N $t (11) = -3.894$ $p< 0.01$,
AP DM N Position 4 $F(2, 13) = 7.795$; $p<0.01$,
DM and N $t (9) = -3.748$ $p<0.01$, AP and N $t (7) = -3.471$ $p<0.01$
AP DM N (small difference) Position 5 $F(2, 22) = 4,391$; $p<0.05$,
AP and N $t (6.395) = -2.102$ $p < 0.1$, AP and DM $t (16) = -2.747$ $p<0.1$

For the most part, the fastest responses are by AP then, DM and then N (graphs 5.1.4, appendix 5), however the differences are small and, for the most part, not significant. In cases where there is a significant difference (at $p < 0.01$), the contributing group is not the AP group. Therefore, they do not need to be treated separately.

[46] There are three Ms for position 2, recollection, and they are spread out. More data is needed for a clear conclusion.

5.5.1.2.3 Brahms

Comparison between DL, M and N

One-way ANOVA shows that there are some click positions for which there is a significant difference but only in the Recollection responses (see also box-plots 5.1.5-6, appendix 5).

Comparison DL, M and N	
Position 1 Recollection	Position 4 Recollection
$F(2, 13) = 5.143; p < 0.05$	$F(2, 21) = 3.398; p < 0.053$

As for the other pieces, these differences are between the DL and M, and N and especially in this case between the DL and N.[47] For position 4, recollection, there are some responses that are 2 bars early (section 5.5.2.3 below). When these are removed the differences between the responses and the original click positions all fall within a quaver. It may be that here, while the DLs could remember exactly the note, the M and N remembered "before the phrase boundary".

Comparison between AP, DM, and N

There is only one position for which ANOVA shows a significant difference between AP, DM and N, and this is at the $p < 0.1$ level: Position 3, Recollection $F(2,18) = 3.064$ $p < 0.1$. The independent sample t-test shows that the significant difference here is between AP and N $t(10.972) = 2.340$ $p < 0.1$. However, as this is only weakly significant and the only response for which this is the case, the results for all the groups are treated together.

Musical Experience Summary

These results indicate that for all positions there are no significant differences between DL and M and, for the majority of cases, there are no significant differences between DL&M and N. Therefore, for most of the discussion below, no distinction will be made between them.

5.5.2 The pieces

Having seen that the results for each excerpt can be treated as one group, it is possible to proceed to the analysis of the responses to the different positions of each piece. The reaction to the positions within each piece were compared in order to investigate whether or not the expected 'u' shape pattern was found implied by the studies discussed in section 5.2.

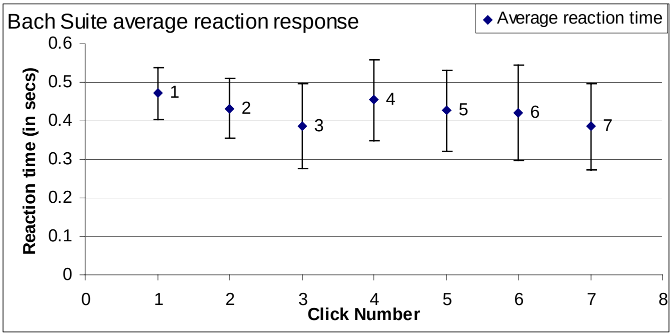
[47] Again there are only 2 responses for M and 3 for N for position 1, recollection.

For each piece, the bars including the click positions are presented, with the positions of the clicks shown with their numbers above the stave. These are followed by the graphs for the 'average' responses for each click position with their standard deviations. The reaction responses are presented in time (sec) relative to the original click positions and the recollection responses are presented in bar proportion relative to the original click positions (box plots for these are given in graphs 5.2, appendix 5).

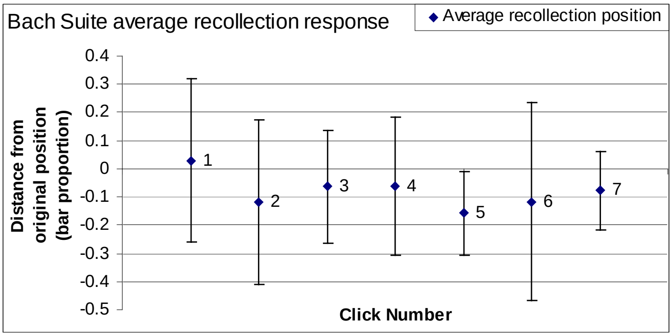
5.5.2.1 Bach Suite

Figure 5.5.2.1.1 Bach Suite, bars 2-3

Graph 5.5.2.1.1 Bach Suite reaction responses



Graph 5.5.2.1.2 Bach Suite recollection responses



Reaction

Graph 5.5.2.1.1 shows that the fastest reactions are to positions 3 and 7, then to positions 2, 5 and 6 and the slowest to positions 4 and 1, though all the responses fall within a semi-quaver after the click. It should be noted that the standard deviation shows that all fall within the same range as each other and a paired sample t-test shows that there are no significant differences.[48]

However, some general characteristics can be identified. The PS note (position 7) has one of the two fastest responses and the two PE notes (positions 5 and 6) have two of the three next fast responses. In addition, these results show that note length is not the only determinant of reaction time as the notes of similar reaction time are not the same length (section 5.2). Otherwise, for example, the response to position 3 would have been faster.

The expectation (section 5.2) is for the reaction time to be fastest at the phrase boundary and more specifically just after the PE (the position of least music-induced cognitive load). In this piece, the PE location is spread and delayed and the first clear signal of the phrase boundary comes with the next PS (position 7). This may be the reason for the fastest response at position 7. The close similarity between reaction times for positions 5 and 6 may be because they are both equally part of the prolonged phrase end.

Recollection

Graph 5.5.2.1.2 shows that all but position 1 are anticipated in the average recollection responses. However, the average recollection responses occur on the same note as the original click for some positions and are anticipated and placed in the previous note for the others. ANOVA showed that there were no significant differences between the responses to any of the positions. In this piece there are so many anchor points at the phrase boundary (the descent to the end, the beginning of the end, the end of the end and the new beginning) and it may be for this reason that the 'migration' does not occur.

Discussion of Bach Suite Results

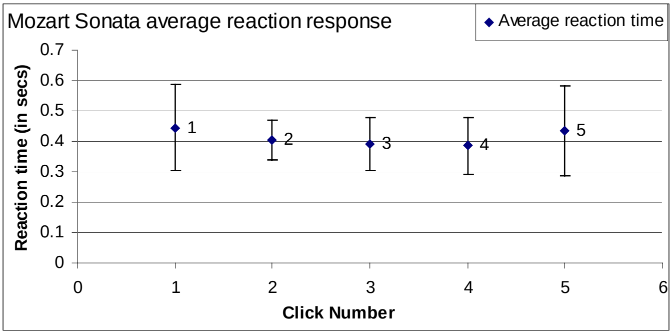
These results indicate that there may be a relation between the position of the note in the phrase and the reaction time and, moreover, that as a result of the prolonged phrase boundary, the clearest boundary point is the third beat of the bar (position 7), with the arrival of the phrase start. That the recollection responses are very accurate may be because there are several anchor points throughout this end of phrase.

[48] As for all of the data in this chapter, box plots were plotted for each of the data sets and before statistical tests were carried out, outliers were removed. However, in order to show the complete picture, the graphs include all of the results.

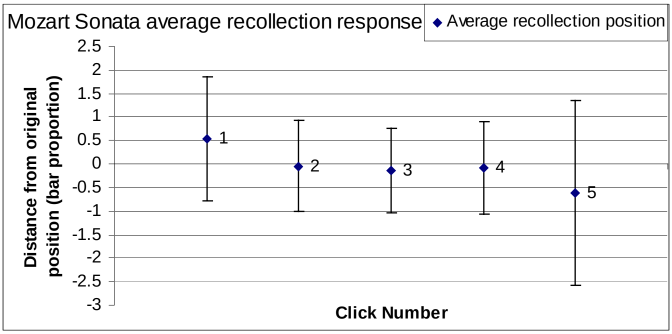
5.5.2.2 Mozart Sonata

Figure 5.5.2.2.1 Mozart Sonata, bar 4

Graph 5.5.2.2.1 Mozart Sonata, Reaction Responses



Graph 5.5.2.2.2 Mozart Sonata, Recollection Responses



Reaction

Graph 5.5.2.2.1 shows a "u" shape with the lowest point on position 4. This could, in theory, be because click position 4 coincides with a long note – thus reducing the 'cognitive load' for that reason alone. However, if this were the case, the response to click position 1 should have the next fastest responses. In fact, the response to position 3 is almost as fast as that to 4, and one of the largest differences is between 1 and 4, both of which are long notes.

Instead, the response pattern follows the theoretical prediction based on the phrase structure (section 5.2). All the responses are within the second semiquaver length after the click and the largest difference of means is between click position 1 and click position 4 (0.06 sec). The difference between groups of responses to clicks is only significant at the level of 0.1 and only for positions 1 and 2 ($t(18)=1.964$; $p<0.1$), positions 1 and 3 ($t(25) = 2.009$; $p<0.1$), and positions 2 and 4 ($t(18) = 1.767$); $p<0.1$).

Recollection

Graph 5.5.2.2.2 shows that positions 2, 3 and 4 are remembered exactly. Position 5 'migrates' back by, on average, 2 semiquavers i.e. back to the phrase boundary, while Position 1 'migrates' forward (i.e. towards the phrase boundary) by about the same amount (not reaching the phrase boundary). For position 1, many (but not all) responses, stay in the same note and so the relatively slow response could be explained by the length of the note. However, at position 4 the note is even longer, and such a delay is not seen. It is therefore possible that the difference is related to the phrase structure and the position of these notes within it. There is only one comparison between groups that shows a significant difference for the recollection responses: 1 and 5 ($t(22) = 3.078$); $p<0.01$). Positions 1 and 2 are different only at $p < 0.1$ ($t(11) = 2.132$; $p<0.1$).

It seems from these results that there is some hint of the expected pattern – the clicks at the phrase boundary 'staying in position' while those before and after not being so stable and moving, in general 'towards' the boundary.

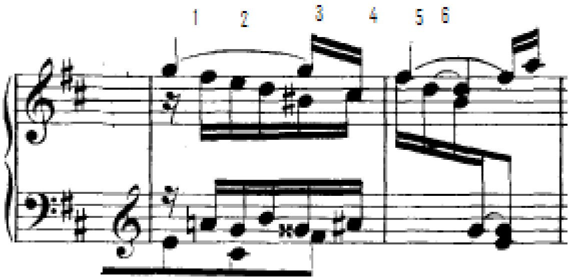
Discussion of Mozart Results

The reaction results follow the prediction of the theory discussed in section 5.2 in that, on average, the nearer the click was to the phrase boundary, the faster the reaction time. However, not all of the differences between reaction times to the different click positions were significant in this respect.

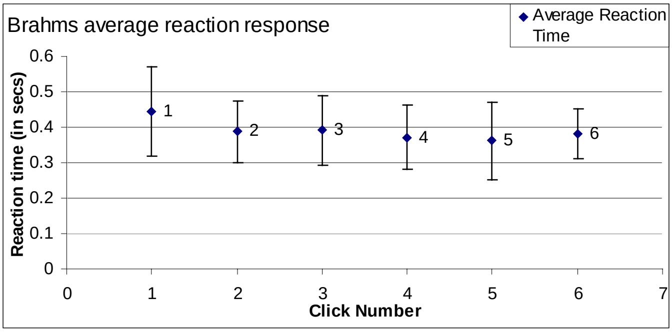
The recollection results also follow the prediction of the theory in that the exact positions of the clicks at the boundary are remembered and those further from the boundary 'migrate' in the listeners' memory though again, not all of the differences between recollection positions for the different click positions were significant.

5.5.2.3 Brahms

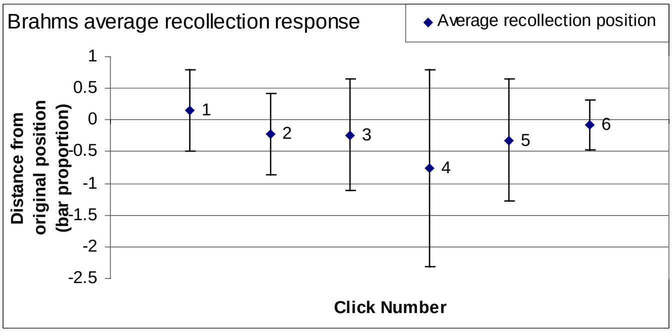
Figure 5.5.2.3.1 Brahms, bars 8-9



Graph 5.5.2.3.1 Brahms Reaction Responses



Graph 5.5.2.3.2 Brahms Recollection Responses



Reaction

Graph 5.5.2.3.1 shows that the average responses, in general, form a "u" shape of decreasing followed by increasing reaction time. The highest point is that furthest from the phrase boundary (but on the bar line) and the lowest is with the first note of the new phrase. Apart from the difference between the first and the second, all the others are within 0.03 seconds. The deviation about the mean for each position is larger than any difference between the positions. The only differences between reaction time between pairs of positions that approach significance even at the 0.1 level, are positions 1 and 4 ($t(14) = 2.024$; $p<0.1$) and positions 1 and 6 ($t(15)=1.833$); $p<0.1$).

Recollection

Graph 5.5.2.3.2 shows that there is a large range of responses relative to the original click positions. Some, especially for position 4, are very far from the original position. The furthest responses to position 4 are two bars earlier (around the start of bar 7). As will be discussed further in chapter 10, the new phrase beginning on bar 9 is 'delayed' by two bars i.e. it could have started on bar 7. It seems, therefore, that the position in the phrase is remembered, just not the correct phrase. This seems to be different kind of click migration from the one encountered in earlier studies (probably because the extracts used previously were much shorter). Here, the migration is to the parallel position (all the listeners that press a key here also identified a PS at bar 5 in the PS task discussed in chapters 3 and 4). This distant migration does not occur for any of the other click positions indicating that this kind of migration only occurs at phrase boundaries and not within the phrase.

Overall, there is a u-shaped graph of mean response positions. The only significant differences are between positions 1 and 2 ($t(9) = 2.489$; $p<0.05$), 1 and 4 ($t(12) = 2.409$; $p<0.05$), and 2 and 4 ($t(3) = 2.705$; $p<0.073$). Removing the responses that were particularly early shows that the deviation remains large. Nonetheless, all the responses are still ahead (to different extents) of where the click would have been: positions 1, and 3-6 are within the preceding semiquaver and position 2 within the preceding quaver (on average). It seems, therefore, that there is not a systematic migration to the phrase boundary.

Discussion of Brahms results

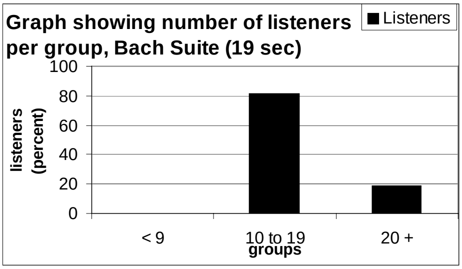
The reaction results suggest that the expected pattern is identifiable but the standard deviation at each position is large. It is interesting that the slowest reaction time is for position 1 which is during the first beat of the bar. This indicates that if the responses are because of the differences in features contributing to 'structural' elements, this response occurs for the phrase boundary but not for the metrical one; otherwise the reaction to position 1 should have been as fast as the responses to position 5 which is on the bar line, but also a phrase boundary.

The recollection responses show that, overall, listeners remembered the location of the click accurately. However, there was migration to the parallel structural position, though only for the position between the phrases indicating a special function of this position.

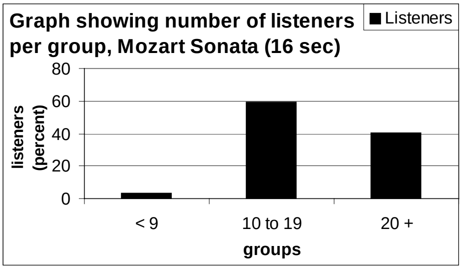
5.5.3 An aside: A distracter task

The distracter tasks were intended to stop the listeners 'listening out' for the click and were secondary to this study (sections 5.2 and 5.3.4). However, one of the questions yielded particularly interesting results. Listeners were asked to estimate the length of the extracts in seconds and gave their verbal response at the end of the second of the pair of extracts. In Graphs 5.5.3.1-3, the responses are presented in three categories: < 9 sec., 10 – 19 sec., and 20 < sec., the actual lengths are given in the graph titles.

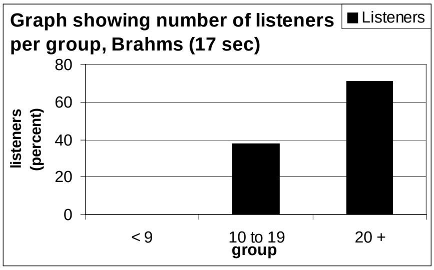
Graph 5.5.3.1



Graph 5.5.3.2



Graph 5.5.3.3



All of the excerpts are actually very similar in length and all fall in the 10 – 19 sec category with the Bach Suite being the longest and the Mozart Sonata the shortest: Bach Suite - 19 sec, Mozart Sonata - 16 sec and Brahms – 17 sec. Many listeners identify this correctly by responding in the 10-19 sec range. However, the proportion of listeners who choose this length changes for every piece, for the Mozart Sonata there is a larger spread with listeners also estimating both < 9 sec and 20 sec <. In the Bach Suite, the vast majority choose the 10 – 19 range, but a small proportion also choose 20sec <. For the Brahms however, the majority choose 20 sec < although the piece is actually two seconds shorter than the Bach Suite. These results follow theories of the relationship between events per units of time (time units being the tactus) and time perception which say that music with less events per tactus are perceived as shorter than those with more events per unit time (Palmer, 1997).[49] These results are for a small number of pieces and listeners, however, not only do they support predictions of other theories, such information may prove useful for a greater understanding of phrase perception.

[49] Thanks to Dr. Justin London for an interesting discussion concerning this matter.

5.6 General Discussion

5.6.1 Reaction

In general, the theory discussed in section 5.2 is supported by the results gathered in the current study; the positions with fastest reaction response are usually those at phrase boundaries and the positions of slowest reaction time are further away.

The position with the fastest reaction time differs slightly among the pieces: for the Mozart it is on the last note of the phrase (position 4), while for the Bach and Brahms it is the first note of the new phrase (positions 5 and 7 respectively). The differences can be explained by the different functions of these notes in the phrase. In the Mozart Sonata, there is preparation for a phrase end, an arrival on it, and a new start. Here the position of least cognitive load is position 4. In the Bach Suite the phrase end is prepared, arrived at and prolonged – the response time at the arrival and prolongation are both the same. The new phrase start has the faster reaction time. It may be that the prolongation means that the cognitive load does not decrease until the new phrase start. In the Brahms there is no 'arrival' on a clear phrase end, instead there is a clear new phrase start and, like in the Bach Suite, it is here (position 5) that there is the fastest response.

The Brahms also included two first beats of bars, one with the phrase start (position 5) and one (the previous one) without (position 1) (see figure 5.5.2.3.1). This allows a comparison between the response to the first beat of the bar with and without the phrase start. In both reaction and recollection, the responses at position 1 were further away from the original click position than those at position 5 (and any other position). The responses indicate that the metrical structure did not affect the responses though further examples are necessary. Different structures around the phrase boundary seem to result in different responses.

5.6.2 Recollection

Only some of the results for the recollection responses are as clearly related to the theories discussed in section 5.2 as the reaction responses. The responses to the Bach Suite do not show systematic 'migration' to one position at the phrase boundary. This may be because the phrase 'boundary' is spread over several beats and there is no single position for which the cognitive load is decreased sufficiently to allow for memory of 'extraneous' noises to be moved to.

The Mozart Sonata has the 'expected' pattern recollection responses with the trend of the click 'migrating' towards the boundary in the listeners' recollections. The responses to the Brahms do not show a systematic 'migration' to the phrase boundary. However, there was, for some listeners, a migration for position 4, the last note of the previous phrase to the equivalent phrase position at bar 5, which did not occur for any other positions. This indicates that the functional/structural position rather than the temporal position is remembered when it is structurally meaningful and not when it is not. This introduces a new idea of ‘migration’ that I have not encountered before in the literature (this may be because shorter examples are usually used in the literature found).

5.6.3 Large deviations, small amount of significant differences, and possible reasons for perceptual deviations

For most of the average values discussed above the deviation of responses is very large and, for the most part, there are no significant differences between responses to different positions. The Bach Suite has the most significant differences between positions. For all pieces, the sample size is very small, in terms of pieces and listeners. To corroborate and build on these results, an increase in sample size is necessary.

The range of reaction times, even for the reaction responses, suggest that even when there is a clear element to react to, in this context the reaction time may be quite long and varied for different listeners. This gives further explanation for the range of responses within each area discussed in chapters 3 and 4.

It seems that explanations for faster response time and more accurate recollection, such as that of long notes (claimed by authors discussed in section 5.2), is not responsible for the trends described. Although there are many confounding factors here it seems that some of the responses and response patterns may be related to the phrase structure and the different characteristics of phrase structures of the pieces. For example, the Bach Suite phrase end is an area that starts with an arrival, continues with a prolongation and resolution and is followed by a phrase start. The phrase end of the Mozart Sonata is prepared, reached and a new start follows immediately. The Brahms phrase end is weak, and the new start, which “could” have occurred two bars earlier, is strong (this is discussed further in chapters 3,4, 10ff). These characteristics of the different phrases may be related to the different response patterns seen here. In addition, the responses here may be related to the different strengths and spreads of responses observed in chapter 3 and discussed further in chapters 10 ff.

Nevertheless, the analysis of listeners’ responses, the differences in response times and the ‘mental distortion’ of the locations of clicks, indicates that the musical features (and therefore phrase structure) are used in such a way as to control the cognitive load and segmentation. These results indicate that this method can be informative not only about the identification of phrase boundaries but also the different phrase part combinations.

Having investigated different aspects of phrase perception of heard music, the next chapter returns to the score based approach of chapter 2 in order to allow comparison with the heard responses.

Chapter 6

Graphic annotation of phrasing:
The second downbeat

6.1 Introduction

Performers usually prepare pieces over a period and have the possibility to analyse the music, look "back and forth" in the piece, and derive their preferred interpretation to which listeners respond (chapter 1). In order to obtain first-hand information on active performers' ideas of phrasing, the following study was carried out. Performers could take as long as they wished to decide on the phrasing, they could see the whole piece at once, they could change their decisions, reporting the final one(s) on the score. The results are analysed in the same way to those of introductory study (chapter 2), in a manner similar to that of the listening studies and are then compared with the latter. Of particular interest is: 1) whether the positions/areas identified are the same or different in the listening and written responses, and 2) what the nature of the differences is, particularly with respect to 'accuracy'.

6.2 Method

Nineteen musicians were asked to annotate phrase arcs on the scores of the Bach Suite, Mozart Sonata and Brahms (the scores are given in appendix 3.1). This part of the study was only possible with active performers; individuals who had learned to read and play music and included only those now play and/or conduct regularly. Only key signature, time signature, bar lines, note length and note pitch, were presented, omitting all other markings such as articulation and dynamics printed on a Sibelius score. The pieces were presented in two formats (different number of bars per page), and in two orders to control for visual cues of formatting and order effects. The musicians were asked to take the pieces, play them through as much as they found necessary and then mark phrase arcs clearly on the music. They were told that if they identified more than one option, they should mark them all (though none did), and that they could provide as many or as few levels of phrasing as they wished. The musicians were asked to return the music within two weeks along with a two questionnaires, one about musical background and the other about the task, to be completed after the musical task (procedure appendix 3.2).

6.3 Results and Discussion

6.3.1 Voices Marked

All of the musicians marked the phrasing over the right hand of the Mozart and Brahms and over the top of the Bach. Two musicians also marked phrasing on the left hand and under the stave of the Bach indicating a 'lower' part within the texture. As the number of markings on the lower parts is so small, however, the rest of this discussion is based only on the markings of the top part.

6.3.2 Effects of presentation (format and order) and musical experience

As summarised in table 6.3.2, no systematic differences were found between the groups of different format and order or with different years of training or playing.

Table 6.3.2: ANOVA and T test results for difference between responses grouped according to musical experience, order of pieces, format of presentation and familiarity with piece

Piece	Years formal training 0-4, 5-9, 10-14, 15-19, 20+ (ANOVA)	Years playing 0-4, 5-9, 10-14, 15-19, 20+ (ANOVA)	Piece Order (T test)	Presen-tation Format (T test)	Familiarity with piece (T test)
Bach	F = 0.20, p = 0.90	F=0.07, p = 0.97	t=-0.152, p = 0.88	t=-0.154, p = 0.88	t=-0.235, p = 0.82, 5 in group
Mozart	F = 0.02, p = 1.00	F=0.23, p = 0.79	t= 0.551, p = 0.58	t =0.31, p = 0.76	t=-0.269, p = 0.79 4 in group
Brahms	F = 1.01, p = 0.39	F = 0.11, p = 0.95	t= -0.22, p = 0.83	t = 0.05, p = 0.96	t=-0.385, p = 0.70 4 in group

Two musicians took part in the earlier listening studies. To check for effects of having heard the pieces in these experimental settings, their results were compared with those of the rest of the group. There were no positions uniquely chosen or omitted by these two musicians. As this group was so small, no statistical test was carried out.

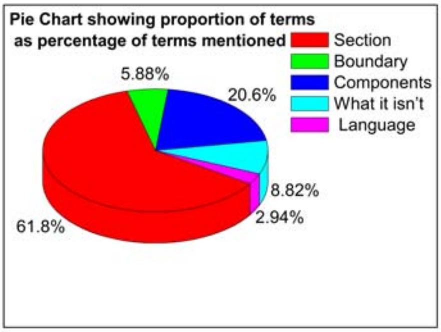
6.3.3 Phrases and sub-phrases

Although the musicians were given the option to provide as many or as few levels of phrasing as they wished, they usually provided only one. In each piece, a small number of musicians did provide more than one level at least once (Mozart, 5 musicians, Bach, 3 musicians, Brahms 4 musicians). Even these musicians however, did not do so all the way through the piece. Moreover, for each piece there are few positions chosen by musicians as 'sub-phrases' that are not chosen by other musicians as 'phrases'. The exceptions are: in the Bach Suite, an additional PS on 4.75, and in the Mozart an additional PE on 3.666 and PS on 3.75. Only in the Brahms is there a longer list - PE: 1.83, 2.83, 3.83, 5.33, 10, 11.83, PS: 1.83, 3, 4.16, 6.16, 8, 10.83, 12.16, 13.16, 14.16, 16.16. These are marked by mainly one musician who wrote in his reasons for marking this phrasing that it makes some standard patterns more 'stimulating'.

6.3.4 Verbal written responses

Like in the listening studies (chapter 3), the musicians were asked to answer two questions about phrasing at the end of the experiment, one general and one specific: 'What, in your view, is the meaning of the term 'musical phrase'?' and 'Please describe what made you put the phrase marks where you did'.

For the first question most of the musicians give a synonym for phrase within the broad category of Section (including, unit, entity segment etc.). A small number also mention the boundary between phrases (breath or pause). Most mention one or more musical features (including harmony, melody and rhythm). Some mention what it differs from (including motif and segment) and some make the linguistic comparison (such as describing the phrase as a means of punctuation). The responses can be grouped according to some of the same categories as in the listening experiment (chapter 3): Section, Boundary, Components, What it isn't, Linguistic comparison. The graph shows the different proportions of terms in the different categories. The number of categories here is smaller than in the listening experiment. Here there are no references to specific difficulties or performance features in answer to these questions (the sample size here is smaller).



These responses indicate that, like in the listening study, the term phrase was meaningful to the musicians both in theory and in practical application. Moreover, though several of the words in the section category are synonyms, the small number of times that each word is used indicates that the musicians are not using an identical definition, indicating that they are not working from a purely theoretical definition.

6.3.5 Pieces

6.3.5.1 Bach Suite

Graph 3.6.2.11, appendix 3.6 shows that, like in the listeners' responses, the start of piece, and bars 3 and 5 are the clearest PSs. Two PS positions are chosen in bar 2: 2 and 2.375. This is one of the few positions for which there is a difference in proportion between the MIDI responses and written responses. For the MIDI there is an almost equal level of response at these positions. In the written version, it is easier to show accurately where the PS is intended and a greater proportion chooses 2.375.

Positions 4.375 and 4.5 are also chosen by a small number of musicians (1 and 4 respectively). One listener starts on the second note of the piece and does the same for bars 5 and 6 (so does another for bar 6). Another listener does so for bar 3.

The PEs for bars 3 and 5 are spread over three beats and the PE for bar 7 over two beats. There is a small spread of responses for bar 2 over the two semiquavers before the bar line and the bar line itself. The majority choose bar 2 itself, but a small number of musicians chose the upbeat. Like the listeners' responses there are different relationships between the PEs and the PSs.

These results show that though the written responses are clearer than those obtained in the listeners studies, there is still a spread of responses over almost all the areas, especially for the PE. This confirms that the reasons for the spread of responses in the listening studies was not purely because of difficulty in pressing keys 'accurately' while listening. Both the areas chosen in the two studies for PSs and PEs, and the spread of responses are very similar.

6.3.5.2 Mozart Sonata

Graph 3.6.3.11, appendix 3.6 shows that the vast majority of PS and PE responses are at 5 positions: the start of piece, and bars 2, 4, 6 and 7. There is a small spread of responses in bar 6 between the first and second semi-quavers of the second beat. There are also a small number of responses at other positions. Some of these are for 'sub-phrases' of the longer phrases but others are included in the main phrases:

There is a small group with several short phrases, their 'extra' PS positions in comparison with the majority are: 3.5 (3 musicians), 5, (1 musician) 5.5 (1 musician) and 7.583 (the same musician). Almost all musicians give PEs and PSs on one note following the next. Only one musician marks an elided phrase (in bar 6).

The areas chosen for PSs and PEs are the same as in the listeners' study and, as in that study, this piece has most between-musicians agreement. Like the listeners' responses there are different relationships between the PEs and the PSs.

6.3.5.3 Brahms

As in the listening studies, graph 3.6.5.11, appendix 3.6 shows that there seem to be two groups of interpretation: those including almost every bar and those including only positions 5, 7 and 9 and possibly bar 13. Also like in the listening experiments, there seem several options as to a location of a possible phrase boundary especially in the second half of the excerpt. The ends of bars 12 and 13 are the most popular PS positions in the second half. For one of the PSs and for more of the PEs there are often two options for the exact location - one note or the next. Like the listeners' responses, there are also different relationships between the PEs and PSs.

6.4 Summary

It seems from these results, that there was high agreement between musicians as to the areas, and sometimes, the exact PS and PE positions.

There were three types of differences between the responses: 1) distinct groups as to the general areas as well as specific positions chosen (as in the Brahms), 2) different groups as to the specific position within a general area (as in the Bach Suite), 3) a small number of areas and positions that were identified by only a very small group or one individual (in all).

These results indicate that the reasons for variety in the responses in the listening study were not limited to the experimental set-up. There are several possibilities, both in terms of general position and location of one note to the next, also when played and annotated.

However, a smaller number of positions were marked here than in the listeners' responses suggesting that both the possibility to read and play and mark at ones own pace rather than having to respond online helps in the identification of the location of PSs and PEs.

To complement this study of musicians "preparing for performance", recorded performances of the case-study pieces are now studied. This is done in order to enable a deeper understanding of the listeners' responses (discussed in chapters 3 and 4) to these performances. A comparison between the two sets of results (performance features and listeners responses) and that between these and the MIDI responses and musical features follows in chapters 10 and 11.

Chapter 7

Performers' Phrasing Study - Performance as communication: Polyphony

'…phrasing tends to dominate performance expression…'
(Friberg and Battel 2002, p. 207)

'He who phrases incorrectly is like a man who does not understand the language he speaks'
(Chopin)

7.1 Introduction

Studies of music performance and perception suggest that phrase structure is one of the central musical elements that contribute to the way pieces are performed, that performers consciously or unconsciously analyse these and other structural elements of the music, that their interpretation of a piece is partly influenced by this analysis and that these structures are reflected and clarified in performance (section 7.2). Phrasing is often described purely in terms of structure-giving features (chapter 1) but performance features are also often described as coinciding with or even providing phrasing. They have also been related to emotional and metaphorical characteristics of music performance (see section 7.2.1 and for example, Meyer 1956; Todd 1985; 1992; 1995 see also chapter 1). Previous studies investigated the relationship between structural elements (such as metrical grouping, phrasing structures or melodic contours) and performance features as tempo and dynamics (section 7.2). In turn, these performance features are thought to help highlight or clarify these structures for the listener.

Aims

To explore patterns of performance features, the locations and degree of changes in these features, and the similarities or differences between performances through study of the literature and analysis of publicly available recorded performances of the case-study pieces. This is intended to prepare for the investigation of the relation between performance features and phrasing, and of how this relates to other musical structures. More specifically it explores the following questions:

1. Are there specific and unique 'phrase-defining' performance features?
2. If so, what are the performance features that highlight phrase structures?
3. Are there differences in performance features among performances?
4. This provides the basis for the investigation of the relationship between performance features and a) listeners' phrase responses to recordings, b) listeners' phrase responses to MIDI renditions, c) written phrase responses provided on the score by musicians, d) phrases identified in music-analytic studies, and e) musical features (chapter 10).

7.2 Previous performance Studies

7.2.1 Studying performances

It is commonly agreed that music is not just the notes on the page, but "the performance". Many factors are involved including performance features such as tempo and dynamic change, use of breath, articulation, use of visual, physical gesture (Parncutt and McPherson, 2002). In this chapter, audio recordings of performances and two performance features: tempo and dynamic change (with some observations about breath) are considered. Performance features may be used for many different reasons and have different effects for the listener. This variability seems to contribute to the continued interest and re-playing of the same pieces. Many authors make a causal connection between structure and, for example, interpretation in general or emotional character (Shaffer 1984; Clarke 1988; Friberg and Battel 2002, p. 199). Shaffer (1984) describes an interpretation as a compact coding of expressive forms, from which an expressive performance can be generated when required.[50]

[50] There are two possible qualifications or additions: 1) The expressive forms that constitute the interpretation are rather abstract; performers use a variety of different expressive strategies to project essentially the same interpretation. This suggests that an interpretation consists of a set of abstract expressive markers that can take a concrete expressive form within any of the parameters available. 2) An interpretation is not only an expressive but also a structural coding. A performer must form an understanding of musical structure, or decide between structural alternatives offered by the music, and encode that in some stable and compact manner. The structural component then acts as a framework around which the expressive markers are organised (Clarke, 1988, pp. 14-15).

Variations in timing and dynamics are notated in the score but these notations are imprecise, have only relatively recently become more detailed as a matter of convention, and are not always those used in performances. In addition, performers rarely write about phrasing of the pieces that they perform. To investigate performance features used in performance, it is therefore necessary to analyse performances and identify the non-notated variations within them.

In most general terms, non-notated variations in timing and dynamics (deviations from the specified notation) can be divided into three main types: *Expressive variations* which are deliberately meaningful or communicative, but not necessarily conscious (Juslin, Friberg et al. 2001-2002). Though it should be borne in mind that even conscious focus of attention may not map simply onto physical parameters that are being varied.[51] *Non-expressive variations* which can be due to technical limitations of the instrument and/or performer, and *random variations* (including imperfections in the perceptual timing and motor system) (Juslin, Friberg et al. 2001-2002). *Expressive variations* can be classified according to their apparent communicative purpose. They may communicate the music's *structure* or express its *character* (emotional or motional).[52]

For Friberg and Battel, a good understanding of structure, theoretical or intuitive, is a prerequisite for a convincing musical performance (2002, p. 199). By applying performance rules concerning such elements as tempo and dynamics in playing, the player enables the listener to interpret the performance. All details of a performance are interpretable by the listener as long as they were derived from performance rules that exploit the listener's previous intra- and extra-musical experience (Sundberg, 1988, p. 66-7).

Sundberg (2000) identified two main principles involved in the communication of musical structure; one aiding categorical perception (which is improved by increasing the difference between categories in performance, such as stretching the frequencies of scale tones or playing short notes even shorter), and the other aiding grouping (by clarifying phrases, metrical units, or harmonic areas with performance features, such as diminuendos at phrase ends). Both involve redundancy: 'This increase efficiency of the musical communication by introducing redundancy; the phrase boundaries are often recognised even without this cue' (Friberg and Battel 2002, p. 212). The importance of redundancy, not only in music perception, is often discussed (for example, Snyder 2000).

---

51 Ian Cross, Personal Communication.

52 Structure and character are not necessarily independent. Character can be seen in terms of *how* the structure is communicated (Friberg and Battel, 2002, p. 212).

7.2.2 Reflection of structure in performance; limiting or emphasising ambiguity

Pieces can have several different structural interpretations, and the primary role of expression is often that of limiting the extent of this ambiguity. However, a performer may also emphasize more conflicting features of the music. Although a performance must aim to be expressively coherent, this does not necessarily entail the resolution of all structural ambiguity (Clarke 1988, p. 15).

Each expressive act projects a particular functional meaning for a given musical structure. This is achieved in a variety of ways, the most general principle being the intensification of gestalt properties of the musical structure that are already evident, or the establishment of gestalt features when the music is structurally neutral (Clarke 1988, p. 15). Examples include: establishment of boundaries in the grouping structure of music by means of changes in dynamic, articulation or timing, imposition or emphasis of a sense of direction towards a structural focal point by means of dynamic, articulation or timing gradients, or modification of the accentual status of events (changes in figure-ground relations) by means of dynamic or agogic emphasis (Clarke 1988, p. 15). In general, the relationship between the expressive aim and means is direct: boundaries are indicated by relatively large parametric changes, directed motion is indicated by graduated parametric increase, and accentual strength is indicated by relative parametric intensity (Clarke 1988, p. 15).

Within timing, dynamics, and articulation, expressive gestures, however, can perform a number of different functions including the indication of a group boundary, a metrical accent, or creating an expressive gradient towards a focal point (Clarke 1988, p. 14). Within at least two of these parameters (timing and articulation), however, the directness of this expressive function is threatened by ambiguity. For example, the lengthening of a note can indicate that it is accented, that it finishes a structural unit at some level, or that the following (delayed) note is of structural importance (Clarke 1988, p. 15). This uncertainty can be clarified in two ways; the sequence in which the gesture appears and its structural context (Clarke 1988, p. 16).

Expressive gestures are, therefore, functionally ambiguous in that they can specify a number of alternative interpretations. These ambiguities are resolved through interactions with underlying musical structure (Clarke 1988, p. 13). In Clarke's study the most expressive changes could be explained on the basis of changes in the position of metrical accents (being played louder, longer or more legato), and group boundaries (discontinuity in the timing, dynamic and articulation curves), a more minor role being the emphasis of melodic peaks (1988, p. 14). Furthermore, the three expressive parameters interact in at least two ways. They may substitute for one another or combine to form expressive complexes that possess a compound function that is not simply the sum of the expressive components (Clarke, 1998, p. 14, see also Shaffer, 1980 and Gabrielsson, 1999 for the study of timing and dynamic in relation to metrical structure and how performers with different levels of musical training play and are perceived).

Under conditions of structural clarity, and for listeners well versed in the musical idiom, expressive characteristics function as responses to, or refinements of, the properties of the music. When the musical structure is weak or indeterminate, however, expressive effects may function primarily to impose a particular structural interpretation onto a neutral structural base (Clarke 1988, p. 17). Referring to investigations by Shaffer, Sloboda and himself, Clarke (1988) proposed generative rules to account for a great deal of the expressive deviations in (piano) performance.

7.2.3 Reflection of phrases in performance

The idea of the phrase is often used in music performance literature though the definitions are either implicit or based on the performance characteristics themselves.

'Musical structure is reflected in physical variables in a number of ways including ritardando and diminuendo at the end of a phrase. The slowing and softening are more pronounced at the end of the phrase and are quite substantial and thus clearly perceptible. The differences in interpretation between the pianists are largely seen in variations within phrases and on a note-to-note level' (Friberg and Battel 2002, p. 202). 'These typical shapes of timing and dynamics are observed in a majority of performances of Romantic music and are important for conveying the basic phrase structure to the listener' (Friberg and Battel 2002, p. 204). Many models were based on 'the idea that musical phrasing has its origin in the kinematic and dynamic variations involved in single motor actions' (Todd 1992, p. 3541).

The degree of change of timing and dynamics reflect phrase level in the hierarchy. For example, for tempo: 'The *ritardando* at the end can communicate the phrase level, with typically a more pronounced *ritardando* at the end of a musical unit of longer duration or at a slower hierarchical level…not only the phrase boundaries but also their hierarchical level – and hence the hierarchical phrase structure of the whole piece – can be communicated, just by changing tempo and dynamics. Similar principles are found in speech, where lengthening is used to communicate phrase and sentence boundaries' (Friberg and Battel 2002, p. 204). Though Friberg and Battel describe this for Romantic music, their first example is from a Classical piece (A sonata by Mozart, 2002, p. 205), indicating that this is not limited to Romantic music.

Friberg and Battel distinguish between different levels of groups: the 'faster level' (small melodic units of a few notes) and 'longer phrases'. At the faster level, grouping (i.e. segmentation) 'tends to be quite ambiguous, often with several possible interpretations' (2002, p. 205), arising from 'contradictory perceptual cues from different aspects of the musical structure, such as the melodic contour or the meter, and can be resolved in performance by inserting a micropause between the last tone of one phrase and the first of the next, which both interrupts the sound and delays the onset of the following tone' (2002, p. 206). 'So communication of this structure can be subject to more individual interpretation than, say, communication of longer phrases' (2002, p. 205). The amount and shape of variation in the phrase can vary between performers (2002, p. 204).

The difficulty of relating performance features and phrasing is illustrated by Friberg and Battel's discussion of the confusion between tension and phrasing. Phrasing tends to dominate performance expression, making it difficult to isolate the more subtle details such as the expression of melodic or harmonic tension (2002, p. 207). 'The most common way to communicate tension seems to be to emphasize notes or areas of relatively high tension, as in the models of harmonic and melodic charge... However, it is difficult to trace the origins of variations of timing and dynamics measured in real performances, since the various tension concepts are often coupled with each other and with the phrasing structure' (2002, p. 207). In the context of tension, Friberg and Battel also mention harmonic characteristics: chords that are more distant from the key are more often found in the middle of phrases, while chords close to the key are more often found in the beginning or in the end of the phrase. Another way of interpreting this is that because the more distant chords are in the middle of the phrase and the ones at the start and end are nearer the tonic, we perceive the phrase-parts in those positions. This is one of the few mentions in these performance studies of such musical features but does not go into detail about where in the chord sequences, for example, different performance features may occur. On one hand, this seems to imply that tension is considered separate from phrasing and should be identifiable as such. On the other, it seems so closely bound up with phrasing that it may not be possible (or desirable) to separate the two ideas. This gives an example of the potential difficulty in relating performance features to musical ones.

Performance features can be studied under experimental conditions with electric pianos that record exact time of note onsets, pressure applied and so on (such as Repp 1995). However, this usually means that these studies are limited to piano music and rely on a small number of performers. An alternative is to analyse publicly available recordings. The measurement of the data obtained from these recordings is less accurate than the first method (though consistency of marking can be checked for by repeated annotations of the same recording) and is restricted to timing and dynamic variations since other measures are more difficult to obtain reliably from sound recordings. However, performances on instruments other than the piano can be studied and several 'accepted', publicly available performances can be compared.

7.2.4 Tempo Change

Studies of timing, usually for piano performance, have identified systematic deviations from strict timing in the performance of experienced keyboard players, and some understanding of the rules governing the timing deviations has begun to emerge (Hartmann 1932; Seashore 1938, pp. 225-253; Todd 1985; Clarke 1988; Sundberg 1988; Repp 1990).

Bengtsson and Gabrielsson, identify four meanings of tempo: 'a) the abstract *mean tempo*, calculated as the total duration of a music section divided by the number of beats in the section, b) the *main tempo*, being the prevailing (and intended) tempo which the initial and final retardations as well as more amorphous caesurae are deleted, (c) *local tempi*, maintained only for short periods but perceptibly differing and (d) *beat rate* … for describing minor fluctuations, which may not be perceptible as such' (Bengtsson and Gabrielsson 1983, p. 50). The 'average' tempo of a phrase may not be evidenced in particular events but listeners find the idea of an average tempo natural and adequate (Gabrielsson 1988, p. 33). For more recent and state-of-the art alternative approaches to the study of tempo in performance see Honing (2005, 2006) and for comparative studies and reviews see Clarke (1999) and Timmers and Honing (2002).

7.2.4.1 Reasons for Tempo change

Several different reasons for tempo changes have been suggested:
1. Structural
2. Expressive
3. Structural resulting in expressive (Clarke 1988),
4. Non-expressive, motor constraints (Penel and Drake 1999), through technical limitations or random variations, (Friberg and Battel 2002).
5. Perceptual: lower-level 'psychological' processes concerning regularity extraction and segmentation into groups (Penel and Drake 1998). This is based on the principle that some inter-onset intervals are perceived as shorter/longer than they are, and are thus performed longer/shorter according to a phenomenon of perceptual compensation. Perceptual biases may result from psycho-acoustic reasons (e.g. high pitch may be perceived as longer than low pitch note) or from processing of events embedded in complex sequences (resulting from grouping into basic units) (Penel and Drake 1998). This is not learnt gradually with acculturation or training but rather a degree of systematicity is fixed in, for example, production or reproduction of simple rhythms. This is not necessarily so relevant for phrasing; when musicians are asked to play 'mechanically', phrase-final lengthening is reduced, while variation related to rhythmic groups is less affected (Penel and Drake 1998).

7.2.4.2 Tempo Change and phrasing

Studies have concluded that tempo changes are particularly important for identifying phrase boundaries using primarily 'phrase final lengthening'.[53] Clarke (1988), for example, summarised three structure-governed principles within the domain of expressive timing in piano performance:
1) Graduated timing changes that indicate grouping of notes, with maxima at group boundaries.[54] When the minimum point is displaced to the left, upbeats predominate and the internal motion of the group is towards its end. When it is displaced to the right, afterbeats predominate and internal motion of the group is dissipation away from its start. The amount of timing modification seems directly related to the structural significance of the segment (Shaffer 1985; Gabrielsson 1988, p. 34).
2) Lengthening of a note inside a group to add emphasis to the following one. This is frequently associated with (1) since the delayed note is often the start of a new group, the previous note being lengthened both for reasons of delay and because it falls at a group boundary. Delay is, therefore, only distinct when it occurs mid-group, or when applied to isolated notes that are not part of a graduated timing curve.
3) Lengthening of structurally significant notes, especially at the starts of groups. Since significant events usually occur towards the start of groups (Lerdahl and Jackendoff 1987), the principle balances the end-effects of (1) (Clarke 1988, pp. 17-19).

The essential principles of expression in (piano) performance, demonstrate three underlying expressive functions:

1) Indicating structural direction through parametric gradients,
2) Indicating group structures through parametric continuities and discontinuities, and
3) Accentuation of individual events through local intensification of contrast (Clarke 1988, p. 21).

These are affected by, for example, the instrument and the performance style (era, ensemble, purpose). For example, some theorists emphasise that there are different patterns of tempo change for different eras. For Friberg and Battel '[i]n music from the Romantic period, large variations in local tempo are an essential part of the performance tradition. Phrases often start slow, speed up in the middle and slow down again towards the last note (Henderson 1936; Repp 1992)' (2002, p. 204). These typical shapes of timing and dynamics are observed in a majority of performances of Romantic Music and are important for conveying the basic phrase structure to the listener. The *ritardando* at the end can communicate the

---

[53] Including Todd, 1985, Shaffer and Todd, 1987, Clarke, 1988, Repp, 1990
[54] The 'group' is not specifically defined here but it is implied that it refers to every level of the 'hierarchical structure'

phrase level, with typically a more pronounced *ritardando* at the end of the musical unit of longer duration or at a 'slower' hierarchical level (2002, p. 204). Phrasing in Baroque music typically involves smaller variations in local tempo than in Romantic music. Baroque music tends to have a more motoric metrical character (as does most contemporary jazz and pop), suggesting the metaphor of a mass moving at a constant speed, creating a kind of musical momentum (2002, p. 204). Difficulties with such general statements become apparent when, for example, considering the French Baroque or even Friberg and Battels own example (section 7.2.3 above).

7.2.4.3 Repetition and Tempo change

'It is often argued that repeated passages should be performed differently in both cases. This is however, not generally confirmed in measurements. On the contrary, there are often striking similarities between the first and second presentation of a thematic group. This is also true for the repetition of a whole piece on different occasions [by the same performer]' (Friberg and Battel 2002, p. 202). Friberg and Battel here make two statements that, they seem to imply, are related to one another. The first, that repeated passages are often similar, later turns out to be based only on timing (p. 202). Interestingly, though the differences are not always statistically significant, the study of the phrasing in songs (chapter 2) indicated that repetitions sometimes differ in terms of their performance timing, both in location and degree of tempo change. This is returned to below (section 7.3.2.2 and chapter 15). Moreover, timing is not the only tool that the performer has. If the structure is being reacted to and displayed in the tempo changes, this does not mean that the other characteristics are the same (such as differences between a statement and its repeat, for example, is the 'echo').

7.2.4.4 Tempo dependency

Expressive transformations accompanying changes in performance tempo are based on structural properties of the music, and can be characterised as the transformation of latent expressive possibilities into manifest expressive features in accordance with the dictates of tempo and musical structure. At faster tempi there are fewer groups in the timing profile than at slower tempi, the additional boundaries being at positions of a structural discontinuity of some sort (Clarke 1988). When the global tempo of a performance is changed, patterns of local timing variations may also change. For example, there may be a tendency toward more expressive timing variation (relative to tempo) at slower tempi (Repp, 1995). The perceptual and motor limits of tone duration may alter the expressive pattern (Friberg and Battel, 2002, p. 202).

7.2.4.5 Limits of the perception of temporal variability

When the degree by which a note is lengthened or shortened in a note or series of notes is too small, the effect cannot be heard. When the quantity is too great, two phenomena have been noted: 1) It is easy to hear but 2) the effect appears musically unacceptable because it sounds exaggerated. The 'correct' or musically useful quantity is between these two. It is typically characterised by the listener noticing the effect but being unable to analyse it correctly in physical terms. What is acceptable from a musical point of view seems to be just beyond what is noticeable (Sundberg, 1988, p. 63). Clarke investigated the detection of small-scale timing changes. Lengthening a certain note by 20-30 ms was detectable in comparison with a strictly metronomic sequence for both short tonal and atonal sequences, whereas about 50 ms was required for detection in a sequence with some rubato in timing (Clarke 1989). 'The quantities by which these rules affect the amplitude and duration of the individual note are sometimes exceedingly small. Yet, the effects thus generated are essential to the impression we get of the performance. This is evident when listening to examples where in the input notation the phrase markers have been replaced by subphrase markers and vice versa. The typical reaction of music listeners is that this simple substitution results in an unacceptable performance of the melody' (Sundberg, 1988, p. 62). This shows that our sensitivity to these minute perturbations of amplitude and duration is very high. If microperturbations appear in the wrong places, they are easy to hear, but if they appear in correct places they are hard to notice (Sundberg, 1988, p. 62). Similarly, in performance, Clarke and Baker-Short found that even in "deadpan" performance, the timing profile still mirrored that of performances with rubato (1986).

7.2.4.6 Tempo Summary

Timing plays an important role in the performance of, and listening to, music and has been studied extensively in the fields of music performance and perception. Timing both of sound and silence is the variable over which the performer has most control, regardless of instrument (Gabrielsson, 1988, p. 29) and it is possible to observe tempo in the study of performances in attempts to compare between pieces for different instruments. Timing is not only related to rhythm but to melody, harmonic progression, single chords and synchrony (Gabrielsson, 1988). Timing is adapted with regard to the global tempo: at different tempi, different structural levels of the music are emphasised and the expressive timing is adapted accordingly showing a close relationship between expressive timing, global tempo, and temporal structure (Honing 2001, p. 50). A large proportion of the timing patterns can be explained in terms of musical structure. These timing patterns help to communicate temporal structure (rhythm, metre, phrase structure) to the listener.

Any performance characteristics have the potential of affecting phrase structure identification. It will be shown here that the largest beat-length changes coincide with phrases described by analysts and with the responses of musicians. However, other changes may also be responded to.

7.2.5 Dynamic Change

7.2.5.1 Identifying dynamic change

Over the centuries, more detailed and systematic dynamic information has been marked in the score, which is now more so than for tempo. An approximately ten-point range is often represented (pppp-ppp-pp-p-mp-mf-f-ff-fff-ffff), which may be related to our perceptual limits (Miller 1956).

A relative nature of the dynamic markings has been identified. For example, a part notated piano (*p*) can be played louder than when followed by pianissimo (*pp*) notation than when preceded by forte (*f*) notation (Namba, Nakamura et al. 1977 reported in; Gabrielsson 1999, p.537).

Furthermore, there seems to be no necessary relationship between the score's dynamic markings and actual performed dynamics, indicating that expression marks in a score tend to be used only as a guide (Todd 1992, p. 3542). Dynamic markings in the score can therefore not be used as a predictor of dynamics used in performance. Instead, the shape of dynamics, like tempo, seems to be a function of structural importance i.e. the more important the boundary the greater the *decrescendo* (Todd 1992, p. 3542).

There seem to be correspondences between dynamics, and tempo and pitch: (1) musical dynamics and tempo change are coupled "the faster the louder, the slower the softer" (especially with reference to Classical and Romantic styles) (Todd 1992, p. 3540-2), and (2) "the higher, the louder" (Sundberg, Friberg & Frydén, 1991).

Tempo change is compared to physical movement and the origin of the pitch relation seems to be physical: wind instruments (including voice) tend to produce louder tones at higher pitches. Often, the most important tone in a phrase is also the highest in pitch. In this case, the high-loud principle produces natural-sounding phrasing (cf. Windsor and Clarke, 1997, Palmer, 1996a, and Krumhansl 1996).

7.2.5.2 Dynamic Change and Phrasing

There seems to be an overriding pattern of dynamic change in phrases: a crescendo followed by a diminuendo. 'Considered as a whole, the amplitude [intensity] profile within each phrase shows an increase toward a maximum at, or close to, the transition from the next last to the last measure and then falls steeply. The termination of each phrase is thus associated with diminishing amplitude' (Gabrielsson, 1987, p. 98, see also Todd, 1992, p. 3542) though the amount and shape of variation can vary between pieces and performers (Todd, 1992, p. 3542, Friberg and Battel, 2002, p. 204).

7.2.5.3 Dynamics Summary

The studies discussed here suggest that phrases are usually, in dynamic terms, shaped by crescendo/diminuendo patterns and that greater changes imply a higher-level phrase. The amount, shape and location of variation over the phrase can vary between performers. However, there are also other musical structural features such as metrical structure that can also influence changes in dynamics.

7.2.6 A Linguistic Connection

Many of the patterns discussed here seem analogous to those found in speech. For example, the emphasis of a note through lengthening is seen as analogous to the lengthening of a syllable in speech (Carlson et al., 1974); in both lengthening is used to communicate boundaries of phrases and sentences (Friberg and Battel, 2002, p. 204), particularly endings. Many languages use 'final lengthening' for signalling a phrase end (Lindblom, 1978). Listening to speech, music listeners have been 'programmed' to interpret lengthening as a possible sign of a termination. However, it seems doubtful that speech is the ultimate source of the musical performance 'principles'. For instance, in some languages, such as Danish, lengthening of syllable duration does not indicate the phrase end but this does not mean that Danish speakers do not lengthen the ends of musical phrases and, conversely, as discussed above, not all lengthening in music indicates termination.

Instead, such couplings as between lengthening and termination are likely to result partly from the listeners' extra-musical experience, including (but not exclusively) speech, and partly from previous musical experiences. The final retard is an example of a coupling of the extra-musical type. It is typically used in motor music, where there is a regular, rapid pulse. Most listeners are likely to associate this pattern of pulses with locomotion. One way to stop locomotion abruptly is by collision, which is unpleasant. It may be for that reason that the final retard is often used in performance of motor music (Kronman and Sundberg, 1987). Todd (1992; 1995) based his model of deceleration towards a phrase end on velocity change in physical motion and Friberg and Sundberg (1999) suggested that slowing towards a phrase end may be linked to the human properties found when adults slow down and stop after running.

7.2.7 Summary of music performance studies

The studies of music performance and its relation to musical structure indicate that:

1. Performance features can coincide with phrase structure.
2. Phrases are indicated by crescendo-diminuendo, accelerando-ritardando (and phrase-final lengthening) (primarily in Romantic music).
3. Higher and lower hierarchical levels of phrases coincide with more or less pronounced changes in tempo and intensity respectively.

4. The global ('average') tempo has an effect on the structural details and therefore may entail changes in local tempi and beat rates.
5. If music is repeated within a piece, the performance features are the same.
6. Often changes in tempo and dynamics also coincide with other elements such as metrical structure (highlighting different positions). It cannot be assumed that changes in these parameters can alone predict phrase characteristics

7.3 Empirical study: Tempo and Dynamic change in different performances of the case-study pieces

The following section prepares for the analysis and comparisons with musical features and listeners' responses in later chapters (particularly chapter 10). Here the methodologies for obtaining the tempo and intensity contours of the different performances of the case-study pieces are described. This is followed by a presentation of the contours and brief descriptions of key similarities and differences between them. In the summary of each section, these are compared to ideas explored in section 7.2.

7.3.1 Method

The present study follows, for example, Hartmann (1932), Povel (1977), Gabrielsson (1987), Repp (1990), by analysing commercial recordings of world-famous artists. The performances examined here reflect skill and interpretative insight at the highest level of performance. Three commercial recordings of the pieces by different performers were analysed and compared (a list of recordings is given in appendix 3.3).

Tempo

The data collection was carried out using the Mustimer program (developed by Murray Allan at Winchester University) which takes as input taps on normal computer keyboard and gives as output the rate taps in beats per minute (BPM) and the time of tap since the first tap (in seconds). In order to have a measure that is comparable across the piece, the unit that was tapped was the 'beat' (tactus). To reduce the effect of tapping error the tapping process for each piece was repeated six times. An ANOVA test for each set of recordings used showed that there was no significant difference between the tapping runs. For example, like for all the others, for the Lipatti recording, the ANOVA showed no significant difference between the runs ($p > 0.99$). An average of at least three of the most similar tapping recordings was taken as representative of each performance recording.

In some cases, there are no note onsets on beats. In these cases, while listening an estimate was made as to the position of the 'missing' beats. When all the data was collected, an average was taken of all the 'estimated beats' from the nearest present note onset until the last estimated beat before the next true note onset.

Dynamics

The performances were each recorded into Praat (Boersma and Weenink), a program originally designed for the study of phonetics which can analyse intensity contours. The intensity contours of each performance were studied individually and then compared to the listeners' responses to the same recordings. They were also recorded (from .wav files) into a program written by Nick Collins[55] (using matlab). Samples were taken at a rate of 100 per second. These were integrated over 200 ms, reflecting a decay constant of 90dB over 200ms after onset (Moore, 1995). The length of integration has a great effect on the resulting contour. To have a more general view, the same data was also plotted using a decay constant of 90dB over 100ms. The graphs presented here are from the 200ms integration time and sometimes show more detail than can be heard directly. Therefore the following discussion concentrates on the larger patterns.

7.3.2 Results

7.3.2.1 Tempo contours

General comparison

The graphs in appendix 3.6 of the tempo contours of the performances of the pieces show several general characteristics including both similarities and differences in:
1) Average tempi between the different performances of the same piece,
2) Degrees of tempo variation in different performances of the same piece,
3) Degrees of tempo variation in different pieces, and
4) Types of tempo variation in different pieces and different performances of the same piece.[56]

A more in-depth discussion of the tempo and intensity contours in relation to listeners', analysts and musicians written phrase identifications as well as musical features is discussed in chapter 10.

In terms of variety between performers, the Bach Suite and Brahms have the largest between-performer differences in underlying tempo while the Mozart Aria has very small differences. Statistically, they are all significantly different:

[55] Many thanks to Nick Collins for much technical help and advice.
[56] 'Degree' here refers to the varying degrees of relative tempo change in general, whilst 'type' refers to contour of the local tempo changes, for example whether they are gradual (over a number of notes with small changes in beat length between each note) or sudden (a sudden lengthening of one note in comparison to its neighbours).

Are the differences between the three tempo contours of each piece significant? (ANOVA)		
Piece	**F**	**p**
Mozart Sonata	20.43	$p < 0.0001$
Mozart Aria	86.77	$p < 0.0001$
Bach Passion	210.0	$p < 0.0001$
Bach Suite	35.03	$p < 0.0001$
Brahms	39.80	$p < 0.0001$
Wagner	14.53	$p < 0.0001$

Furthermore, the Brahms has the biggest difference in average tempi among performances and the greatest within-performance range. The Mozart Sonata and Aria have the smallest range and the smallest within-performer tempo changes.

Average beat length and standard deviation of each tempo contour (sections 7.2.4 and 7.2.4.4.)			
Piece	**Performer**	**Mean beat length (sec)**	**Standard Deviation**
Bach Passion	Leonhardt	0.55	0.15
	Furtwängler	1.04	0.23
	Cleobury	0.68	0.12
Bach Suite	Gendron	0.67	0.15
	Rostropovich	0.93	0.19
	Casals	0.80	0.17
Mozart Sonata	Uchida	0.89	0.14
	Lipatti	0.75	0.08
	Brendel	0.77	0.11
Mozart Aria	Böhm	0.38	0.03
	Solti	0.36	0.03
	Ostman	0.40	0.03
Brahms	Kovacevich	1.20	0.36
	Gould	0.60	0.25
	Lupu	0.92	0.35
Wagner	De Waart	1.04	0.29
	Barenboim	1.05	0.19
	Böhm	0.91	0.27

Comparison of tempo contours of the different performances

In the Bach Passion there are different types of note lengthening: lengthening of individual notes and gradual lengthening of several notes. Several different types of musical elements seem to result from these changes in tempo, including accentuation of particular notes (such as the g bar 3.125) and the starts of bars, and lengthening of some (orchestral and vocal) phrase ends respectively. There does not seem to be a systematic link between vocal phrases and tempo changes though there seems to be some coincidence between listeners' responses and changes in tempo (chapter 10). The only clear gradual, reduction in tempo in all performances is at the end of the movement where there is a clear ritardando to the end. In general, the differences in tempo within each performance seem relatively small. However, they do differ between the performances.

In the Bach Suite, the three tempo contours are very different from each other. Gendron lengthens the last note of bars 1, 2 and 4. Casals does the same but lengthens even more the first beat of bar 2 and also the upbeat to bar 6 as well as having a more gradual ritardando in the beats preceding bar 5. Rostropovich has a ritardando during the end of bar 2, culminating on the last note. He then has a relatively short first note of bar 2 and lengthens most of the bar. This is followed by a series of fast-slow notes, the slowest being on 3.875, gradually getting faster towards the end of the extract. This is unusual, usually a change in tempo, if there is one, is a ritardando rather than accelerando, such as in the cases of phrase-final lengthening or ritardandi towards the end of sections.

In the Mozart Sonata, three positions are lengthened (bars 2, 4 and 7) by all three performers. Bar 6 is lengthened by Lipatti and Uchida but not by Brendel.

In the Mozart Aria the tempo contours, like those of the Sonata, are very similar to one another and there is a very small range of tempo variation. The Ostman contour has the most similar tempo contours for the first and second halves of the extracts and the difference not statistically significant for any performance.

In the Brahms the last quaver beat of bar 8, has the most dramatic note-lengthening in all the performances. The tempo contour for Kovacevich shows a regular, relatively smaller lengthening of for example the first beats of bars 3, 5, 7 and 11. The Kovacevich recording is much slower than Gould's. The tempo contour for Gould's recording shows fewer peaks while Lupus's has least tempo variation.

In the Wagner, there are several positions of note lengthening at which all three performers coincide including the last beats of bars 4 and 9, bars 26-27 and 35. There are also several positions where the tempo contours are different, including bars 13-14, 23 and 29-30.

Tempo Contours Summary

The tempo contours show that there are different types of tempo changes. Some are gradual (such as Brahms, bars 15-16) and others are more local note lengthenings (such as Wagner bar 9). In all of the pieces there are areas for which all performances have tempo change. There are many areas in which tempo changes only in some recordings, or where there is a difference in degree of change between recordings.

Friberg and Battel suggest that Baroque music has smaller variations in local tempo than Romantic music (section 7.2.4.2). A comparison of the tempo contours shows that the greatest deviations from the 'main'[57] tempo are at the end of the excerpts; It is only in the final ritardando of each extract that there is a greater difference between the Brahms and Wagner on one hand and the Bach extracts on the other, contributing to the larger standard deviation in the former. If these ends are not included in the comparison, then the tempo variations are similar between the two Bach extracts and the Brahms and Wagner. These results indicate that tempo variations in these Baroque pieces are not always much smaller than those of these 'Romantic' ones.

Characterising performances of Romantic music, Friberg and Battel highlighted the importance of the *ritardando* in communicating the phrase level, with, typically, a more pronounced *ritardando* at the end of the musical unit of longer duration or at a 'slower' hierarchical level (Friberg and Battel, 2002, p. 204). As will be explored further in chapter 10 (Brahms) the coincidence of listeners' responses, analysts markings, written responses with the performances seem to further support this description. However, in this piece, other music structures are also highlighted with changing tempo. For example, in the Kovacevich the first notes of most bars are lengthened, usually with little preparation or continued lengthening (chapter 10).

Clarke's statement that at faster tempi there are relatively fewer groups in the timing profile and at slower tempi relatively more (Clarke 1988) is supported by, for example, the Brahms tempo profiles (especially the comparison between Gould and Kovacevich). Repp states that when the global tempo of a performance is changed, patterns of local timing variations may also change. For example, there may be a tendency toward more expressive timing variation (relative to tempo) at slower tempi (section 7.2.4.4). This also seems to be supported by the Brahms, the two Bach extracts and even the Mozart Sonata. However, in the rest of the case-study pieces, even when the mean tempo (Bengtsson and Gabrielsson 1983) is very similar (Mozart Aria) or when there is not one performance that is systematically faster or slower than the rest (Wagner) there can be differences in the degree of 'expressive timing' and the number of phrases (sections 7.2.1. and 7.2.3)

7.3.2.2 Intensity contours

In general, the intensity contours (appendix 3.6) show similarities and differences in average intensity, intensity range, and locations of intensity changes.

[57] Bengtsson and Gabrielsson's meaning b) above section 7.2.4.

Average intensity and standard deviation for each intensity contour			
Piece	**Performer**	**Mean Intensity (dB)**	**Standard Deviation**
Bach Passion	Leonhardt	93.38	1.64
	Furtwängler	69.75	6.13
	Cleobury	89.7	2.06
Bach Suite	Gendron	73.51	4.06
	Rostropovich	70.83	3.99
	Casals	95.9	4.25
Mozart Sonata	Uchida	82.64	4.00
	Lipatti	92.45	4.80
	Brendel	90.70	3.90
Mozart Aria	Böhm	97.83	0.87
	Solti	93.83	1.21
	Ostman	94.13	0.73
Brahms	Kovacevich	76.05	4.24
	Gould	86.3	3.55
	Lupu	81.27	3.73
Wagner	De Waart	76.7	2.62
	Barenboim	75.4	2.58
	Böhm	74.1	2.51

Bach Passion

The intensity contours (graphs 3.6.1.6-8, appendix 3.6.1) of the different performances are similar; the peaks coincide with the singers' high notes and the troughs with their rests. The phrase structure of the singers' and orchestral parts do not coincide all the time and this was identified by listeners (chapter 10). However, the vocal part dominates much of the intensity contours so it seems that such intensity contours that express the whole texture at once do not fully represent the information heard by the listeners. The effect of loss of information is less pronounced in the Mozart Aria as there was less contradiction between voice and accompaniment.

Bach Suite

The intensity contours (graphs 3.6.2.8-10, appendix 3.6.2) seem more different than for the other pieces, such as the Mozart Sonata. Overall, the Gendron intensity contour descends steadily across the whole section and more dramatically at the end, while the Rostropovich contour has a high start (with the first chord) and, for the most part, remains within the same range. However, great extremes in the Rostropovich intensity contour compared with relatively smaller peaks and troughs in the Gendron result in a large difference between the intensity ranges (Gendron, about 24 dB and Rostropovich about 49 dB). The Casals intensity contour seems to be similar to that of Rostropovich staying mainly in the same range. However, there is a generally lower section around bar 3 with the extreme minimum at bar 3.25. Unlike the other two recordings, the minimum at bar 2 is very shallow. There seems to be more preparation for the phrase end at bar 3 but generally the troughs are in the same positions as the other recordings. However, some characteristics, such as the minima of bars 3 and 5 are shared among all the performers.

Mozart Sonata

The three intensity contours (graphs 3.6.3.8-10, appendix 3.6.3) are similar in that two have a similar intensity range (about 41.4 dB - Lipatti, 43.7 dB - Uchida) and the third has a slightly larger range (about 46.3 dB - Brendel), and all three have three areas of minima on bars 2, 4 and the end of the excerpt. For each of these the lowest trough preceded by a relatively low minimum at the start of the bar. There are also minima at other positions such as the start of bar 5 in all three recordings, and bar 1 of Uchida and Lipatti. They both coincide with long notes, which, as the piece is played on the piano, fade anyway. However, the attack of the long note seems quieter too. This shows that minima identified in raw intensity contours cannot be used as sole predictors of phrase boundaries.

Between each minimum, all the contours have a rise and fall of intensity, two smaller ones for the first two phrases and a larger one for the last which is most pronounced in the Brendel and 'interrupted' in the other recordings (especially the Uchida).

However there are some differences. The most noticeable is that there is a minimum on bar 6.333 (first semi-quaver) of the Uchida recording but only a relatively weak one in the corresponding part of the Lipatti and Brendel recordings. In contrast, the tempo contours, have almost identical tempo changes at this position. The Lipatti recording is also a little faster in underlying tempo than the Uchida recording, which emphasises the gradual changes in volume more than Lipatti.

Mozart Aria

The intensity ranges for the Böhm and Solti performances are the most similar for all the performances studied here. Both have a range of about 34 dB. The Ostman has a slightly larger range of 43 dB. The minima areas in the Solti contour are also in Böhm contour. However, there are more minima in the latter. The Solti contour is smoother, with less breaths and more gradual and long diminuendi while the Böhm contour is more chopped with less prepared minima i.e. less diminuendi. The Ostman intensity contour is very similar to that of Böhm with minima in the same positions. There are also a small number of 'extra' minima in the Ostman (bars 3.5, 5.25-5, 7.25, and 13), though most of these are relatively shallow (graphs 3.6.4.8-10, appendix 3.6.4).

Friberg and Battels' (2002, p. 202, section 7.2.4.3) observe that phrases are repeated in the same way. The Solti example shows that, although overall the positions highlighted remain similar, there are subtle yet important differences between the repetitions. For example, the first 8 bars are more clearly subdivided than the second set of 8 bars. Moreover, there is a much larger 'breath' at 4 than at the parallel position (bar 12) (see also chapters 2 and 15).

Brahms

In general, the three intensity contours are similar (graphs 3.6.5.8-10, appendix 3.6.5). Lupu's and Gould's ranges are similar (41 and 40 dB respectively), while the Kovacevich has a smaller range (33 dB). All gradually rise to around bar 8 and then generally fall, with a second, smaller rise around bar 13. As in many of the other pieces, the general positions of least intensity are also similar. The general intensity contours of the three performers seem more similar to each other than the respective tempo contours (graph 3.6.5.7, appendix).

All three performances have the largest decrease in intensity over a short time during the last notes of bar 8. For Kovacevich and Gould there is also a trough on bar 9.25. However, there seems to be a difference between the troughs at these two positions – the first (at the end of bar 8) is the end of a longer diminuendo, while the second results from a 'lengthening' of the accentuated first notes of bar 9 which die away on the piano, creating another trough.

All performers have minima at the ends of bars 1, 8, 13, and 15 and the whole of 16. Kovacevich has two more positions of minima - the parallel positions of bars 2, and 10. The intensity contours of Gould's bars 5-7 and 7-9 (with 9-11 of lower intensity overall) are more symmetrical than the respective positions in the other performances. There are a number of other differences in the locations of minima: Gould and Kovacevich on bar 3 while Lupu at the end of 3 (and Kovacevich small trough at end of 3), Gould end of 4, to 5 while Kovacevich and Lupu end of 5, Gould and Lupu only, middle of 12, Gould middle of 11 while Kovacevich on 11 and Lupu end of 11.

Wagner

The intensity contours of the Barenboim and De Waart recordings are different. However, the minima are mainly in the same bars e.g. 4-5, 9-10, 15, 17, 22-23, 26-27, 32 and 35. Some are, however, of different lengths among performances. The intensity contour of Böhm's recording is similar to that of Barenboim, with generally more extreme minima (graphs 3.6.6.6-8, appendix 3.6.6).

In De Waart's recording, the troughs are longer and are often prepared by a gradual decrease. This can be interpreted as indicating a deliberate diminuendo or 'silence' (accentuated diminuendo). In Barenboim's recording, the troughs are shorter and sharper, maybe indicating a physically necessary breath. For example, in bar 12.75 (c-f) there is a wide trough in De Waart's contour but not in Barenboim's. The following bar is of relatively low intensity in both. This indicates that the change in direction that occurs here is clarified more in De Waart's than in Barenboim's performance.

In Barenboim's recording, there are gradual rises and falls in intensity, with only some of the positions of low intensity prepared by a gradual decrease or by a longer lasting trough. For the louder sections of the Barenboim, it seems that the greater use of vibrato causes a faster rate of increase and decrease of intensity in the higher intensity ranges of the Barenboim than the De Waart. It seems that the Barenboim disguises the positions of breath that are there only by physical necessity and accentuates only some, while the De Waart emphasises more of the breaths regardless of location. The results of the tempo and intensity contours of the Brahms and Wagner in particular seem to support Clarke's statement that when the musical structure is weak or indeterminate expressive effects may function primarily to impose a particular structural interpretation onto a neutral structural base (1988, p. 17, section 7.2.2).

These results indicate that the intensity contours may be used to distinguish between breaths (of necessity), which are short and preparing diminuendo, and rests (of phrasing) and may therefore be used for phrase identification in performances.

Intensity Contours Summary

The above observations indicate that in each piece there are a number of positions with similar relative intensity levels (peaks and troughs), which are the same among the performances. Around some of these positions there is a similar degree intensity change (crescendo/diminuendo), while for others the intensity changes are different. In particular, in some performances, the change in intensity (particularly the diminuendo) is more gradual than in others.

In most cases, these intensity changes also coincide with changes in tempo (diminuendo with ritardando). However, there are some cases in which they do not. This indicates that dynamic change may influence the listener responses, even without the support of tempo change. This implies that the general correspondences between dynamics, and tempo "the faster the louder, the slower the softer" (Todd 1992, p. 3540), though being common, does not always hold (section 7.2.5.1). With the same kinds of exceptions the idea that the end of each phrase is generally associated with diminishing intensity is supported by most of the results here (Gabrielsson, 1987, p. 98, section 7.2.5). One technical exception is for the pieces for solo voice and accompaniment; in cases where there is a dominating line in recording (such as the Furtwängler, for which the intensity of the solo is much higher than the accompaniment), ambiguity is created by different structures in melody and accompaniment that is indiscernible in the intensity contours.

In the current study, the changes in the different parameters cannot be externally and individually controlled neither can the degree of change in tempo and intensity be changed. However, listeners' responses indicate that sometimes very small changes in intensity and in tempo result in listeners' responses (section 7.2.4.5 and chapter 10).

According to Friberg and Battel the differences in interpretation between pianists are largely seen in variations within phrases and on a note-to-note level (section 7.2.3). The results of the two case-study pieces for piano (and the other case-study pieces) indicate that, though there are differences within the 'phrase', there are also some differences at some phrase ends and starts. For example, some performers highlight bars 2 and 6 in the Mozart Sonata (both with changes in tempo and in intensity) while others do not. Similarly, in the Brahms, though bars 9 and 16 are highlighted in all the contours, there are other phrase ends and starts highlighted by some performers and not by others both in terms of tempo and intensity changes. For both pieces, these differences coincide with differences in responses to performances (chapter 10).

7.4 General summary

This study suggests that in terms of position and degree of change in tempo and intensity, there may be generally great similarity between performances (such as the first two phrase ends of the Mozart Sonata) but also differences (such as in the Bach Suite and Brahms). Moreover, for each piece, there are some positions with similar contours, and others with more different patterns in different performances.

The study suggested that the distinctions often mentioned between phrasing in music of different eras (section 7.2), is not always helpful (for example, the Bach Suite). It further suggests that there are some subtleties that should be borne in mind when using such contours. Between and within the contours there are different types of tempo and intensity changes: some are gradual (lasting several beats), some are sudden, some are large and some are very small. The results for the Wagner indicate that the use of breath seems to change depending not only on piece but also in performance. The intensity contours of Bach Passion indicate that these contours should be treated with caution in the case of polyphonic music. One part may dominate intensity contours recorded even though a single part may not dominate during listening. There may be other performance features (such as changes in timbre and articulation) that play a role in phrasing but were not investigated here.

In general, the patterns described in the literature discussed in section 7.2 were also identified here. The relation between these patterns and listeners' phrase responses are discussed and the reasons for similarity and difference in tempo and intensity change within and between performances will be directly related to musical features in chapter 10.

Chapter 8

Previous theories and the current listening study results: Cacophony

8.1 Introduction

A general review of the main studies relevant to the subject of phrasing and phrase perception was presented in chapter 1. This was followed by a general discussion of the results of the listening and performance studies (chapters 2-7). Now, before embarking on a detailed discussion of the case-study pieces and a closer analysis of the results of the listening and performance studies (chapters 9-13), several previous studies that have made at least some reference to phrasing or phrase boundaries will be discussed in detail. These discussions end with the application of the respective theory and implementation rules presented, or comparison of experimental results in this study, to the ones obtained in each study.

The main aims of these discussions are:

1) To review in detail the current definitions of, and assumptions about, the term 'phrase'.
2) To identify the cues and explain decisions on phrase identification and definition.
3) To investigate the results of these studies in light of their underlying theories and evaluate their general applicability.
4) To interpret results of the current study using the various theories and rules given in these studies.

The studies discussed here are experimental, theoretical and computational (mainly rule based but some have an element of memory-base). The music-analytic approaches will be discussed in chapter 9. Almost all are related in some way to Gestalt theories (chapter 1) and several are even more inter-related. Most of the studies refer directly to the term 'phrase', however, even those that concentrate specifically this, explain that they do not have a comprehensive theory of phrasing and often, they limit their theory *a priori* to certain types of musical genres or textures.

The following approaches will be examined as they each give a different perspective and represent the essence of the respective theories:

a. Lerdahl and Jackendoff (1987) primarily use musical instantiations of Gestalt-based ideas and present a rule base that identifies groups, one level of which they call the phrase. As with the rest of their approach, their theories of grouping are not only often referred to in, but also strongly influence, much subsequent work.
b. One example of this development is the experimental work by Deliège (1997), which was followed by a study based on a more general interpretation of the Gestalt principle of 'change' (Deliège 1998). This study also provides some musical analysis and carries out a listener experiment using one of the case-study pieces used in the current study (Wagner). Therefore, the discussion here concentrates on the 1998 paper and a comparison with the results obtained for the Wagner in the current study.
c. Cambouropoulos (2001; 2003) develops two independent but potentially complementary theories of boundary detection. The first relates to Gestalt principles of change and proximity, and to both Lerdahl and Jackendoff's and Deliège's theories. The second is also related to a Gestalt principle, this time the principle of similarity. It further develops part of Lerdahl and Jackendoff's theory and also relates to that of Deliège's.
d. Bod (2001; 2002) presents an approach to melodic phrase structure that aims to identify phrases in monophonic folk songs. He uses the Essen folksong collection which has 6251 pieces already annotated by an experienced musician to learn the different pitch and rhythm patterns of each individual phrase in his training set and calculates the probability of finding these patterns as phrases in new pieces. This is then tested new folksongs from the same collection and the new annotations are compared to those of the experienced musician. Bod claims that his model shows that there are cases for which Gestalt principles and music-theoretic harmonic and metric norms are not useful and that in these cases the memory-base is more successful in annotating these pieces.
e. Ferrand et al. (2002) develop a probabilistic model of listeners' melodic segmentation. They use cues that are also used in the studies mentioned above, but this time in a probabilistic manner in which entropy change is taken as indicating a boundary. They compare their results to those of Deliège's experimental study (1998) based on the same piece by Wagner.
f. Temperley (2001) develops a rule base that uses a preference system, that is similar to that of Lerdahl and Jackendoff in its general approach, but is more specifically directed towards phrasing. There are several additional differences between the two approaches: in Temperley's approach there are fewer rules and musical cues, some of which are different from those used by Lerdahl and Jackendoff. Moreover, rather than being a purely rule based approach, one of

the three rules is based on a memory-based approach of learning the average phrase-length of pieces in an annotated corpus.

g. Palmer and Krumhansl (1987a) carry out a listening study that assesses phrase-completeness ratings by musicians and tests the contribution of 'pitch' and 'temporal' cues to these ratings. They use two different approaches to model the responses: one by Lerdahl and Jackendoff, though rather than using the grouping theory, they use their metrical and time-span theories, and the other being Krumhansl and Kessler's (1982) pitch 'fittedness' profile. Palmer and Krumhansl present a second experiment in the same paper in which they shift the coincidence of pitch and temporal information. However, in the current study, only their first listening experiment is discussed because it already provides many relevant observations and questions.

The discussion of each study begins with a brief summary concentrating on the parts that are most comparable to the current one. This is followed, in most cases, by a comparison between my own results of the listening study and their results and/or theories. This is revealing because it allows an evaluation and exploration of the wider applicability of the theories or characteristics. Most of the rules and cues discussed seem to be based on 'score' features rather than purely 'performance' ones. Therefore, all of the comparisons are made using only the responses to MIDI. The exception is the discussion of Deliège's study, which is based on listeners' responses to performances.

8.2 Lerdahl and Jackendoff's Grouping Structure Preference Rules

8.2.1 Lerdahl and Jackendoff's theory and rule base

8.2.1.1 Introduction: Groups

Lerdahl and Jackendoff describe a theory, based primarily on Gestalt principles and also drawing on Schenker's and linguistics theories, that enables the identification of 'possible structures' that can be assigned to a piece that correspond to listeners' grouping intuitions (Lerdahl and Jackendoff 1987, chapter 3).[58] Their theory is represented in a series of Grouping Well-Formedness Rules (GWFRs) and Grouping Preference Rules (GPRs) for grouping structure. 'When hearing a piece, the listener naturally organises the sound signals into units such as motives, themes, phrases, periods, theme-groups, sections and the piece itself. Performers try to breath (or phrase) between rather than within units. Our generic term for these units is *group*' (p. 12). They investigate monophonic music or polyphonic as monophonic music (p. 37).

8.2.1.2 Grouping Well-Formedness Rules (GWFRs)

The GWFRs establish the formal structure of grouping patterns and represent what the listener brings to the perception of music (p. 39). Lerdahl and Jackendoff propose that grouping is a hierarchical property and in their GWFRs they outline the formal conditions for hierarchical structure. The rules define a group by stating the conditions that all possible grouping structures must satisfy. These conditions define a strict, non-overlapping, recursive hierarchy (p. 37).

8.2.1.3 Grouping Preference Rules (GPRs)

Coupled with the GWFRs are GPRs which establish which of the formally possible structures can be assigned to a piece correspond to listeners' intuitions. They give 'relative preferences among a number of logically possible analyses' (p. 42). The hypothesis is that one hears a musical surface in terms of the analysis (or analyses) that represent the highest degree of overall preference when all preferences are considered (p. 42). The aim is to express analytically the relations

[58] In the rest of this chapter, references to this text are given just with page numbers.

that the listener intuitively perceives, i.e. the unconscious principles of the perceptual organisation. For most of the rules, Lerdahl and Jackendoff use a sliding window of four notes (the GWFRs and GPRs, pp. 345-347, are given in appendix 8.2.1).

Lerdahl and Jackendoff point out that, sometimes, thematic considerations require the start of a new group where local detail and cadential considerations strongly favour the continuation of a previous group. They leave a formal characterisation of these phenomena that will involve grouping structure, metrical structure, metrical irregularities and time-span reduction for future research (p. 62).

The GPRs are closely tied to the Gestalt principles. With the exception of the GPR 1 (excluding groups containing a single sound event), two principles Proximity (2a, slur-rest rule, 2b, attack point rule), and Similarity (change rules – 3a register, 3b dynamics, 3c articulation, 3d length, and timbre) underlie the rules. If both principles reinforce each other, the resulting grouping intuition is strong. If there is conflict, the resulting intuition is ambiguous. One principle can override the other (p. 41). The proximity rules describe the length differences and the change rules describe the modifications in the acoustic or temporal state of sound structures, in relation to Gestalt theory (further discussion of the relation between the GPRs and Gestalt theory is given in appendix 8.2).

8.2.1.4 Summary of Lerdahl and Jackendoff's musical cues

The GPRs concentrate on several musical cues, all of which can be seen as musical interpretations of Gestalt principles:
- Temporal gaps: temporary changes in length - long notes or rests (GPR 2).
- Longer term changes in note length (GPR 3d).
- Pitch gaps: pitch jumps (GPR 3a).
- Changes in dynamics or articulation (GPR 3b and 3c).

Having presented these basic cues for 'low level grouping', Lerdahl and Jackendoff present rules for the groups themselves and how to connect the smaller groups to form larger ones (GPR 1; that groups should not be too small, may be included in this type of rule). '[R]elatively pronounced' occurrences of the features in GPRs 2 and 3 are taken to signal more important grouping boundaries (GPR 4). Groups should be put together in an 'ideal' way, i.e. with the imposition of an external template in which there are two parts of equal length (GPR5). Any repetition of material should occur in parallel parts of the groups. This has implications for the location of group boundaries (GPR6). Grouping structures should result in more stable time-span and/or prolongational reductions (GPR 7). It seems that repetition and harmonic stability are features that are applicable only after an initial low-level grouping.

8.2.1.5 Long notes and rests, phrases and bars

In Lerdahl and Jackendoff's approach, long notes and rests signal the end or boundary between groups. Conversely, long notes and rests are seen as accented units that signal the start of a metrical unit (Longuet-Higgins and Steedman 1971, Republished 1987; Steedman 1977; Longuet-Higgins and Lee 1982; Longuet-Higgins and Lee 1984 and even Lerdahl and Jackendoff's MPR5, p. 84; Spiro 2002). It should be borne in mind that, as with many of the cues discussed in this work, these observations are specific to western music listeners. Cross-cultural studies indicate that the importance attributed to relatively long notes as the main beat by western listeners, is not shared by those from other cultures (Stobart and Cross 2000, p. 89).

In the context of western music, the use of long-notes in both metrical and phrasing theories signals a contradiction: on one hand, long notes are treated as indicating a unit's start by the relative accent in the metrical context. On the other, they signal the boundary between units, by the 'gap' they create. In practice, this can be the case; often the last note of a phrase is relatively long and does fall on a down-beat and in a way this brings out one of the differences between metrical and phrase structure.

8.2.1.6 The performer's influence on preferred hearing

For Lerdahl and Jackendoff, the performer, choosing an interpretation, is deciding how he hears the piece and how he wants it heard. An interpretation includes a (largely unconscious) preferred analysis of the piece with respect to the grammatical dimensions addressed by their theory. As grouping structure is a crucial link between the musical surface and the more abstract time-span and prolongational reductions (pp. 124-178), the perception of grouping is one of the more important variables the performer can manipulate in projecting a particular conception of a piece (p. 63).

The principal influence the performer has on grouping perception is in the execution of local details affecting the choice of small-level grouping boundaries, through GPRs 2 and 3, and larger boundaries, through GPR 4, for example through subtle variations in articulation. However, they explain that the performer's conscious awareness of these strategies often does not go beyond "phrasing it this way rather than that way"; in large part these strategies are learned and used unconsciously. In making explicit the effect of such strategies on musical cognition, Lerdahl and Jackendoff explain that they have suggested how their theory potentially addresses issues relevant to performance problems.

8.2.1.7 The listener

Lerdahl and Jackendoff are describing 'the "musical intuitions of the experienced listener", the '*listener who is experienced in a musical idiom*' (p. 1), who does not necessarily have a conscious grasp of musical structure; 'an acculturated listener need never have studied music' (p. 4). Indeed, for them, the 'concept of the "experienced listener" is no more than a convenient delineation. Furthermore: 'Occasionally we will refer to the intuitions of a less sophisticated listener, who uses the same principles as the experienced listener in organising his hearing of music, but in a more limited way. In dealing with especially complex artistic issues, we will sometimes elevate the experienced listener to the status of a "perfect" listener – that privileged being whom the great composers and theorists presumably aspire to address' (p. 3) thus assuming a difference between listeners' responses and mainly treating the experienced ones.

8.2.1.8 An example of previous experiments testing the GPRs

Deliège carried out experiments testing Lerdahl and Jackendoff's theory asking if: 1) musicians' and non-musicians' segmentations follow the rules' predictions, 2) the rules cover all grouping situations in music, and 3) they are of equal perceptual salience (Deliège 1987). She gave musicians and non-musicians music from Baroque, Classical, Romantic or early twentieth century repertoires to 'segment'.

According to Deliège, her results showed the rules' validity. However, musicians responded significantly more in accordance with the rules than non-musicians (within the musicians, the degree of musical training did not exert a strong influence on the results), in line with Lerdahl and Jackendoff's notion of 'experienced listener' (Deliège 1987, p. 356). However, segmentation into subgroups did demand musical training and as long as the player's performance was not crucial, non-musicians can segment, for the most part, in accordance with the rules. The GPRs might therefore be considered to apply broadly after all (Deliège 1987, p. 356).

Furthermore, subjects' segmentations suggested new rules according to, for example changes in harmony or instrumentation and/or sound density symmetry given by the repetition of pairs of identical sounds, and change in direction of the melodic contour. Moreover, a study using other musical sequences and experimental conditions might reveal new factors (Deliège 1987, p. 357) and the results showed some difficulties with the length rules (Deliège 1987). According to Deliège the sensitivity to the 'gap' in music perception may be considered a key element in grouping. The results suggest two distinct mechanisms: for sound duration, and for acoustic qualities.

Deliège found that some rules, such as change in timbre and proximity of attack point, seem more preferably applied. This does not mean that less preferred rules are "bad" (Deliège 1987, p. 357).

8.2.2 Analysis of the test pieces using Lerdahl and Jackendoff's Grouping Preference Rules (GPRs) and comparison of listening study results with outcome of this analysis.

8.2.2.1 Application of GPRs and presentation of results

To explore the extent to which the GPRs relate to the results of the current study, the GPRs were applied to the case-study pieces and the results were compared to the listeners' responses to MIDI renditions.

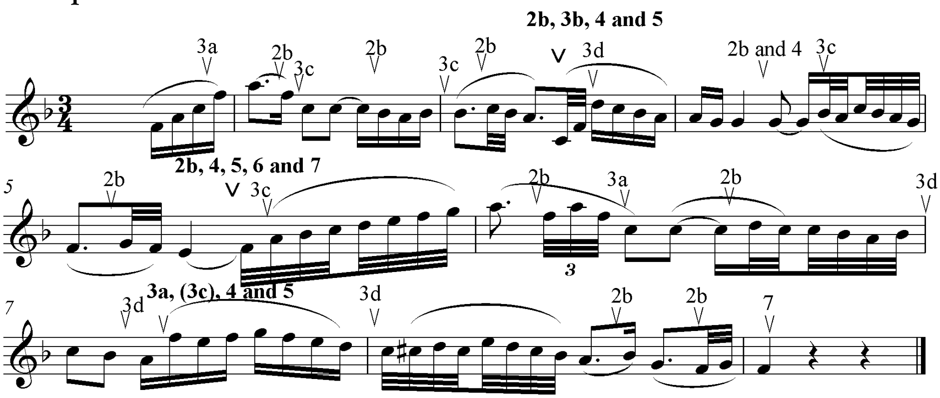
It is not always completely clear how the GPRs should be applied. In some cases, the methodology for identifying features has not been fully developed (such as in the parallelism rule) while in others there are many options and the 'preference' is not clear. As a result different structures may be identified by applying the same rules. This is not necessarily an 'incorrect' outcome; sometimes there several options are also perceived by listeners.

The GPRs were developed for monophonic (or homophonic) music (p. 37) so the rules were applied only to the melody line. The results of the application of the GPRs to the case-study pieces are presented in musical form for the first example. They are also presented in graphic form for comparison with the results of the listening study. The MIDI renditions are the nearest comparison from this study available for all of the case-study pieces, with the written music on which the GPRs are based. Therefore, only the responses to the MIDI renditions are discussed. The performance features that are not accounted for by the GPRs were not available to the listeners. The dynamics and articulation information (of rules 3b and 3c) and the Slur part of the Slur/Rest rule (rule 2a) were unavailable to the listeners and are therefore not considered here. Lerdahl and Jackendoff refer to phrase boundaries rather than phrase starts and ends. Therefore, in this section only the phrase starts are referred to.

8.2.2.2 Results and Analysis

The pieces

Example 8.2.2.2.1 Mozart Sonata



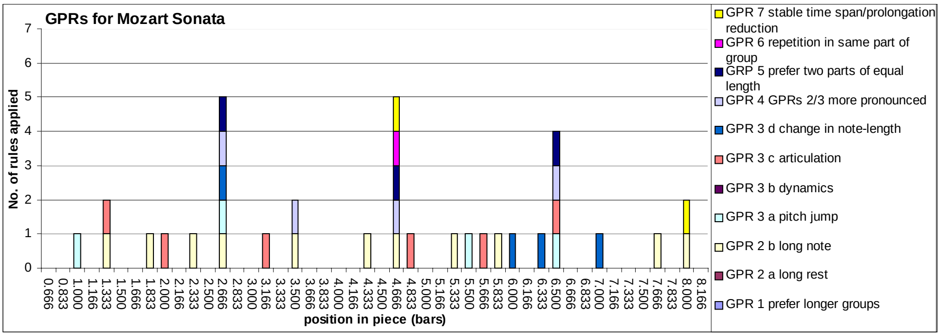
Rule application

Example 8.5.1.2 shows that there are many positions at which one rule can be applied, one position where two can be applied and three positions that have a relatively high coincidence of rules. For the latter positions, there is one where both low-level types apply (bar 2) and one each where one or the other low-level rule types apply (bb. 5 and 7). It seems, therefore, that neither rule-type can be ruled out though both occur often throughout the rest of the piece, indicating that neither happens only in conjunction with higher-level rules.

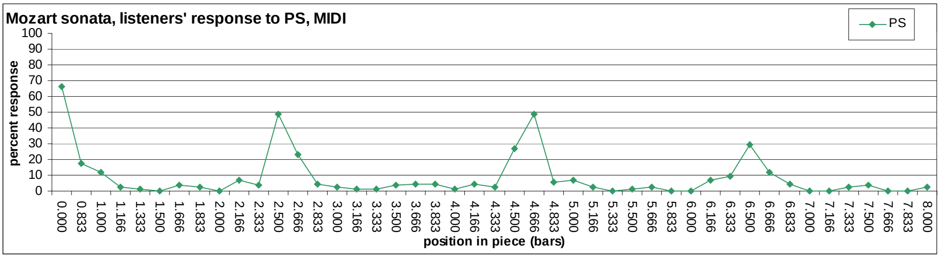
Comparison with MIDI listening experiment results

The following graphs show the locations of the successful GPRs (graph 8.2.2.2.1) and the listeners responses to MIDI (graph 8.2.2.2.2).

Graph 8.2.2.2.1



Graph 8.2.2.2.2



The main peaks of the GPR application coincide with those of the responses (bars 2, 4, and 6).

In terms of the proximity rules, rule 2a does not occur. For rule 2b, of the 8 long notes that occur on their own, 5 coincide with a response. There are 4 more occurrences of long notes: bars 2, 3, 4, and 8 where they are combined with one (3 and 8) or more (2 and 4) other rules. In bar 3, the other rule is GPR 4 and coincides with few responses and in bar 8 the other is GPR7, which relates to the phrase end and the end of the excerpt. Therefore, it is not surprising that there is no phrase start response here. In the other two positions, 2b coincides with both 'low level' rules of pitch jumps and other temporal changes and 'higher' level rules of symmetry, parallelism, stability of time span prolongation or intensification. The response is high around these positions.

The change rules 3a and 3d occur at 7 positions. Five are alone (bars 1, 5, 6 and 6.333, and 7). Of these, two do not coincide with a response (bb. 7 and 6), 1 coincides with a small response (bar 5) and the rest coincide with higher responses. These two remaining positions (bars 1 and 6) are very close to other rules (1 is near the beginning of the piece and 6.333 is the last note before one of the peaks). However, it is not necessarily connected just because of the proximity; the rule would suggest a boundary before this note and not after it. Again, the remaining positions coincide with several other rules and it is only here that we see a high response amongst listeners.

Coincidence of GPRs 2 and 3 occurs only in bar 2, forming one of the highest peaks both of rules and of responses.

'Higher-level' rules can only be identified at positions identified already by GPRs 2 and 3. The higher-level rule of bar 8 has been discussed. At the other three positions at which the higher-level rules succeed there is a high response by listeners.

Therefore, on the basis of GPRs 2 and 3, some positions can be identified. These positions form a basis for the application of the rest of the rules and lead to the identification of the main boundaries, which coincide with the areas identified by listeners. The spread of the results and the relationship between this and the rules is discussed below. Further analyses are in appendix 8.2.2 and summaries are given here.

Summary of Positions identification

For the Mozart Aria and Bach Passion, as for the Mozart Sonata, although the responses are highest around the positions indicated by the peaks in the application of Lerdahl and Jackendoff's rules, responses are spread over a greater area. For the Bach Passion, the areas end at the position identified by the rules.

For the Bach Suite the areas of highest response coincide with peaks of GPRs. However, there are also a number of positions of responses, even for positions having a small number of rules (and limited to those of rule 2).

For the Brahms the number of GPRs applicable seems to be small compared to the other pieces. This piece is more polyphonic than the others; although there is a distinction between the 'melody' and the 'accompaniment', this is not as clear on listening as in the other pieces. This may be why a relatively small number of rules are successful in this piece. This may also explain the reduced coincidence between responses and positions identified by the rules of Lerdahl and Jackendoff, e.g. bar 5. This is not surprising as the GPRs rules were developed for monophonic or homophonic music.

In a more extreme way than the Brahms and unlike the other pieces, in this Wagner there are many peaks in the response that do not coincide with any of the positions identified by the application of Lerdahl and Jackendoff's rules. At the same time, there are many positions identified by the application of Lerdahl and Jackendoff's rules that do not coincide with a response. There are also relatively few positions where the higher level rules are clearly applicable.

More generally, the 'lower-level' rules in GPRs 2 and 3 identify instantaneous 'boundary' features: the 'experienced listener' (p. 1, section 8.2.4 above), on hearing a feature, is represented as knowing that a group has ended and the next is beginning.

The 'higher-level' rules rely on a comparison with what has come before and what will follow: GPR 6 relies on memory and a test of similarity so that the 'boundary' is identified in retrospect (after the repetition has begun). However, it is implied that the exact position of the beginning of the repetition, and therefore group, is identified. GPRs 1 and 4 rely on the identification of the relative importance of different features and adding more weight to the more 'important' features, thereby identifying the more important grouping positions and the positions of the longer phrases. GPR 5 relies on the identification of the groups that are most similar in length assuming an importance of identifying regularity in group-length. GPR 7 relies on identifying the harmonically most stable parts of the group and favouring these as group boundaries. These rules therefore rely on memory of what has come before and identifying relative importance. However, all this is done on the basis of choosing between positions already identified with GPRs 2 and 3.

8.2.3 Discussion of the comparison of the MIDI responses with Lerdahl and Jackendoff's theory and rules

The application of the GPRs to the case-study pieces

The application of Lerdahl and Jackendoff's rules (appendix 8.2) showed that, in all but the Brahms, there are three levels of rule-coincidence: individual, two rules, or more. In the Brahms there are only two levels: individual and two rules. Two rules coinciding often results from the identification of similar features using different rules rather than the coincidence of different features. The only position-determining rules are those in GPRs 2 and 3 and these are applied first. After this, the 'higher level' rules (GPRs 1 and 4-7) are applied but only to the positions already identified. There are relatively few positions on which many low-level or high-level GPRs coincide.

Lerdahl and Jackendoff use the term phrase in association with groups and indicate that they are higher-level groups, but do not exactly define the relationship between the terms. It seems that the comparison between the MIDI results and the application of the GPRs can shed light on the relationship between the perception of phrase boundaries as identified by listeners and groups identified by the application of Lerdahl and Jackendoff's model of perception.

Comparison of the MIDI responses with the application of the GPRs

The comparison of the MIDI responses with the results of the GPRs application shows that in almost all the cases, the boundaries identified by MIDI listeners coincide with the boundaries identified by GPRs. In many cases, the highest MIDI responses coincide only with peaks of the GPRs i.e. a one-to-one correspondence in most cases and in the general area of MIDI response. This suggests that the GPRs predict not only grouping boundaries, but that the highest peaks i.e. when several GPRs coincide indicate phrase boundaries, indicating that the GPRs can be used to predict phrases.

However, not all the individual GPR rules coincide with high MIDI responses all the time. There is, as suggested by Lerdahl and Jackendoff, a distinction between 'lower' and 'higher' level rules. The rules that combine to make the peaks within the GPRs are the higher level ones: GPRs 1, and 4-7. These rely on at least some comparison with what has come before or what will follow. However, in comparison to GPRS 2 and 3 these are the less clearly defined rules and rely more on the theorist's interpretation of them. The peaks identified in this study using these rules usually coincide with high MIDI responses. It seems that these are the most important features for phrase boundary identification.

There therefore seems to be a correspondence between the phrases identified in the listening study and groups identified by the interpretations and application of the GPRs. The results suggest that the higher-level rules are applicable to phrase identification, while the lower-level ones, though they automatically coincide with the phrase positions identified as a result of the order of rule application, seem not to indicate only phrase starts.

The above discussion revealed the features of the GPRs that correspond with phrase boundaries identified in the MIDI responses. However, there are some discrepancies between the two.

Lerdahl and Jackendoff's Group boundaries and the Spread of responses

One of the main discrepancies between the peaks of MIDI responses and peaks of GPRs is that while the peaks resulting from the GPRs occur at individual positions (instantaneous), the MIDI responses are usually over a longer span, the size of which changes from position to position (see also chapters 3-5 and 10). The spread of MIDI responses may be for three reasons:

1) Reaction time varies among listeners. If this were the main reason for the spread, the spread would be the same regardless of musical context. However, the spread varies among positions, indicating that musical context is important.

2) The task description and the way it was carried out by the listeners. Listeners were asked to press a key at the beginning of a phrase, lift their finger at the end and press a key again at the beginning of the next phrase. Listeners sometimes lift their fingers where they thought the end of the phrase was going to be. Others waited until confirmation of the end or start. The variety results in a spread of responses.

Lerdahl and Jackendoff's examples and descriptions indicate that, for them, the grouping beginnings and endings are conflated to one boundary that occurs between specific notes. The schematic representation, both in terms of temporal definition and in terms of graphic representation, is instantaneous rather than over an area. The results of the current study (chapters 3, 4, 10 and 11) suggest phrase ends and starts can be indicated over an area. Comparison with the phrase boundaries identified using the GPRs shows that this area usually precedes the phrase boundary according to them, but rarely trails behind.

3) A reflection of the effect that different musical characteristics have on expectations and identifications of the ends and beginnings of phrases. As described above, listeners' responses for phrase starts often coincide with rests or long notes (see section 8.2.2.2). However, in contrast to the implication Lerdahl and Jackendoff's model, the responses do not always coincide only with the end, but rather the beginning of, or during, the rest or long note. This occurs even on

first listening, when the listener does not yet know what will follow the position taken as the end of the phrase (for example, the Bach Suite and Bach Passion). It seems, therefore, that features other than those exactly at the phrase boundary may be contributing to the identification of these positions. In Lerdahl and Jackendoff's model the majority of the rules concern these exact positions rather than areas. Although there is a difference in distribution pattern of responses between musicians and non-musicians in boundary areas for some examples, there is a spread of responses for both groups in all pieces analysed.

There are several examples of musical contexts in which even if the above reasons for the spread are taken into account, the idea of a boundary between notes or the exact location of the phrase start is problematic. The MIDI responses for the pieces that have upbeats and phrase starts on weak beats or weak parts of the beat show that the majority of listeners identify the beginning of the phrase at the start of the upbeat or weak beat. However, there are some listeners who identify the phrase start on the nearest strong part of the beat or strong beat or somewhere between the two (Mozart Sonata and Bach Passion). This suggests that although metrical structure does not seem to be equivalent to phrase structure, it does affect the perceived position of phrase starts.

Application of GPRs to polyphonic music

The results of the GPRs seem to most closely coincide with listener responses for pieces that had a melody and accompaniment or monophonic pieces. As may be expected, the results of the rule application coincide less with the MIDI responses for the 'polyphonic' music in which the 'melody' is less clear (especially the Brahms). The reduced success for the polyphonic pieces is mainly because the gap features are either absent or are overshadowed by other features.

In general, the GPRs therefore seem to be useful for music that is predominantly monophonic or homophonic and even for music that has a melody and accompaniment while the application to more polyphonic music is problematic.

Groups or grouping boundaries

Lerdahl and Jackendoff's rules concentrate on identifying grouping boundaries and so their rules are based on boundary-specific cues. The rules allow for precise and immediate identification of a boundary as precisely as between two notes (or if there is a rest between two notes, at the end of the rest).

The GPRs broadly indicate two hierarchical levels of grouping for which different rules apply. However, there does not seem to be a precise distinction between the two.

There are two main categories of low-level grouping rules (GPRs 2 and 3). For both, groups are identified primarily in terms of their boundaries. The only ways in which the 'content' of the groups can be referred to is as a 'non-boundary' at that level, or that there is a lower-level group boundary which is subsumed in the higher- level grouping.

In terms of the higher-level groupings (GPRs (1), 4-7), the ideas concerning the relationships between groups are: 1) relative importance - how pronounced the cues are, 2) the repetition of material should occur in the same parts of the groups, 3) the lower-level groupings should follow the ideal symmetrical structures of two groups of equal length, and 4) that they have reached a position of relative harmonic stability. In all but the third (which uses an imposition of a template, presumably learnt from previous musical experience), the model uses information heard elsewhere in the piece.

The above ideas about the constitution of higher-level groups entail the boundaries between groups and some external limitations on their proportion. However, there do not seem to be limits on length of, or proportion between, groups identifiable from the specific piece being processed. The only cue for selecting higher-level boundaries that could affect length of groups is the final rule which looks for stable time-span and or prolongational reductions.

8.2.4 Summary

The comparison of the results of the MIDI listening study with those of the application of the GPRs suggests that these cues indeed contribute to the perception of group (and phrase) boundaries. The approach models the reaction to cues instantaneously and in retrospect resulting in the identification of group (and phrase) boundaries.

The above comparison indicates that most of the higher level units identified by the GPRs are identified in the MIDI listening study, implying that high level group boundaries are equivalent to the phrase boundaries in the majority of the case-study pieces. However, the spread of responses in the MIDI listening study suggests that in order to tackle the idea of phrasing, it is essential to consider also their internal structure. This includes the use of information within phrases, to predict, and expect, future events. This is, to some extent touched on in the GPRs, specifically through the preference for groups that have stable time-span and/or prolongational reductions, which requires the knowledge of the internal structure of the group.

8.3 Deliège's Segmentation study

8.3.1 Deliège's underlying theories and aims
8.3.2 Deliège's method in relation to the current one
8.3.3 Results of the current study in relation to Deliège's results considering the differences in methodologies
8.3.4 Summary of Deliège's results and findings relevant to the current work

8.3.1 Deliège's listening experiment and underlying theories

Deliège carried out an experiment which used the same cor anglais solo from Wagner's *Tristan und Isolde* (Deliège, 1998) as is used in the current study (chapters 3 and 10).

Deliège investigates the effects of training and familiarity with the piece on segmentation and memory of it. She included a similar task to the phrasing one in the current study, with groups of musicians and non-musicians. This work is relevant to the current discussion because it is based on a model of mental representation of musical form during real-time listening and, moreover, investigates segmentation. Though segmentation is not necessarily the same as phrasing, the principles Deliège uses to describe segmentation have been previously used for phrasing (for example, Temperley, 2001). She explains that her model is primarily based on Gestalt principles of group formation, especially on similarity and proximity, which are at the basis of Lerdahl and Jackendoff's work (chapter 8.2). Deliège refers to her previous experimental studies which, she says, confirmed the importance of Lerdahl and Jackendoff's principles of similarity and proximity in grouping (Deliège, 1987). She refers to rhythmic grouping implying either a distinction between rhythmic groups and others or that all groups are rhythmic. The reasons for, and consequences of, defining the groups as rhythmic are not developed in Deliège (1998).

Deliège reminds us of the possible generality of information perception regardless of mode, both in terms of Gestalt principles (originally developed for visual information processing), and in terms of 'cue abstraction' (first developed for language processing). She also draws on theories of organisation of space or 'cognitive maps' (Tolman, 1948) and on the categorisation and classification of structures (Rosch, 1975). These search for cues derived from exact or varied repetitions of material. Rosch describes three levels, cue identification, identification of variants, and grouping (Deliège, 1998, p. 64).[59]

[59] Further references to this paper in section 9.3 are given in page number only.

8.3.1.1 "Même et Different"

For Deliège, the main grouping mechanisms are governed by the Gestalt principles of 'Même' and 'Different'. Deliège's previous experiments illustrated the principle of 'Même' - as long as the same basic cues are perceived, structures are grouped together – and 'Different' – perception of contrast (new cues) leads to perceived separation (Deliège, 1989, Deliège and El Ahmadi, 1990). According to Deliège, the major articulations – phrases, périodes, sections – are thus delimited to form a mental schema of the heard work (p. 65). Deliège contrasts this segmentation with the 'mental line', which is concerned more with memory of the piece than active segmentation on first hearing (p. 65, see also appendix 8.3).

8.3.1.2 Deliège's phrases

Deliège mentions the term phrase three times, each time indicating that at least some segmentations are phrases. She does not state a definition of the term phrase (nor of segment). The first time, she describes the cues leading to 'major articulations' such as phrases, periods and sections. This is done in the context of the main features – those leading to perception of 'Même' and 'Different' (p. 65).

The second time, she describes two phrases identified by listeners (bb. 1-4 and bb. 5-9) and implies that bb. 10-14 also constitute a phrase (p. 67). This clarifies the distinction between phrases and periods and indicates that, in her view, the main features are those contributing to the perception of 'Même' and 'Different'. These opening bars (as in many pieces) initially seem to be the clearest section of the piece. However, Deliège's results in the Mental Line experiments cast doubt on this representation of the perceived phrases of this section (see also appendix 8.3).

Deliège gives a specific example in which she concentrates on bars 21.75 and 22.75 (positions VIII and IX below). She explains that these two positions show a clear difference between musicians and non-musicians (p. 83). Non-musicians choose both positions VIII and IX while musicians choose VIII at the expense of IX. The non-musicians' response is possibly induced by the ascent of a fourth, which follows a passage of rhythmic-melodic repetition. At this position, the leap of the fourth can leave an impression of tension, which awaits the final point of the phrase end, driving lay listeners to choose both VIII and IX. Musicians, on the other hand, choose VIII over IX because of its stronger cadential role.[60] She attributes this to their greater musical training; they are more sensitive to the syntax. Here Deliège first specifies that Lerdahl and Jackendoff's *Slur and Rest* rule (1983, pp. 43-5, chapter 8.2 appendix 8.1) is the most relevant (p. 83).

---

[60] Musicians choose IX as a secondary segmentation point in the 'hierarchical' segmentation task (p. 68).

8.3.1.3 Deliège's Experimental Method

Four subject groups - non-musicians (undergraduate students (NMS) and post-graduates (UR)) and musicians (students from the Music Academies and Conservatoires (MS) and teachers from the Music Academies and Conservatoires as well as a group of professional musicians (PM)) - listened to a cor anglais recording of the piece. Listeners were given three tasks: a simple segmentation task (identifying one level), a hierarchical segmentation task (identifying three levels) and a mental line task. To test consistency, Deliège compares the response to the simple segmentation task to that of the hierarchical segmentation task (p. 68). The following discussion concentrates on the first task, which is most similar to the current study.

Subjects were instructed to listen to the piece as if it were a text, discourse or story, with a view to introducing segmentations, in a manner analogous to punctuation in a written text (p. 68). All subjects had one familiarisation listening in which they could watch a screen that had a time marking on it and write the times of their punctuations.

Each group was divided into two (except the PM). The first group had one familiarisation listening (NMS1, MS1, UR1, PM) and the second had three before they were asked to respond. This discussion concentrates on the first as it is most similar to the current study.

8.3.1.4 Deliège's results

Twenty-one potential segmentation points were identified. These are represented in roman numerals in appendix 8.3. Deliège's method of recording the performed note durations and location of subjects' responses seems to indicate that each of the possible points refers to one note (pp. 68-9).

Quotation A

- I [end of b. 4] et II [end of b. 9], soit avant et après la sequence contrastante;
- III [middle of b. 15], soit à la fin de la reprise variée du motif initial;
- IX [end of b. 22], soit avant l'usage insistant du motif rythmé introduit par un triolet;
- XI [beginning of b. 27] et XVIII [end of b. 36], soit avant et après le motif en battues

  D'autres segmentations devraient intervenir, mais au niveau des indices récoltés dans l'étude évoquée, elles ne sont pas prévisibles (Deliége, 1998, p. 67).[61]

[61] I [end of b. 4] et II [end of b. 9], are before and after contrasting 'sequences'; III [middle of b. 15], is at the end of a variant of the first motif; IX [end of b. 22], is before the use of the repeated rhythmic motif introduced by a triplet; XI [beginning of b. 27] et XVIII [end

Deliège explains that as a result of Deliège (1991 and 1992b), the following parts of the piece are chosen as salient:

Quotation B
'Les sujets musicians avaient été frappes par la phrase intiale (mes. 1-4 [I]) et son retour varié en fin de sequence (mes. 10-14 [after II to a bar before III]), après le passage d'une phrase contrastante (mes. 5-9 [I – II]). L'allure rhythmée de la tête de ce motif contrastant, introduit par un triolet de sons conjoints descendants, et l'exploitation qui en est faite ultérieurement dans la mélodie est soulignée, ainsi que les formules en battues qui débutent à la mesure 27 [XI]. Parmi les non-musiciences, seuls quatre sujets sur quinze semblent avoir été (explicitement du moins) frappes par ces aspects, mais, en revanche, tous on souligné l'effet "cor de chasse" entendu vers la fin (mes. 29-35) [XIII and XIV – XVII]' (p. 67).[62]

According to Deliège, the results from her previous studies (Quotation B) allow her to predict the six positions (and their respective reasons, Quotation A) according to 'indice' (indicators which are musical characteristics of the piece) that follow the 'Même' and 'Different' gestalt rules. However, three of the positions are not mentioned in the summary of previous results: positions IX (the related VIII, section 8.3.1.2) and position XVIII. Furthermore, the results are presented in terms of bars rather than exact note positions. This is particularly important in the case of bars 10-14 where there is a relatively close roman numeral (in bar 15, III) but the reference is presumably not intended for that position. As a result, the relationship between the important musical elements of the piece and the predicted segmentation points does not seem as clear as it is presented in 'Moyens et perspectives' (p. 67).

Overall, Deliège's results show that there are 10 positions in the piece (I, II, III, VIII, X, XI, XV, XVIII, XIX and XXI) in which more than 50% of listeners identify a segmentation point. Of the 6 positions that Deliège explains according to the 'Même et Different' gestalt rules only position IX does not reach a 50% response.

It seems to remain unclear what the reasons are for the relatively high response to the 5 positions and the other 15 positions not predicted by the Même et Different principle indicating that the explanations are complex and therefore interesting;

---

of b. 36], are before and after the rhythmical motif; The other segmentations should 'come up', but on the level of cues in the study mentioned, they are not foreseeable.

[62] The musician subjects were struck by the first phrase (bb. 1-4) and its varied return at the end of the 'sequence' (bb. 10-14), after the passing of a contrasting phrase (bb. 5-9). The rhythmic nature of the start of this contrasting motif, introduced by a triplet of slurred descending notes and the use that is ultimately made of it in the melody is underlined, as are the rhythmic forms which start in bar 27. Among the non-musicians only four out of 15 subjects seem to have (explicitly the least) been struck by these aspects, but on the other hand, all have emphasized the effect of the 'hunting horn' heard towards the end (bb. 29-35).

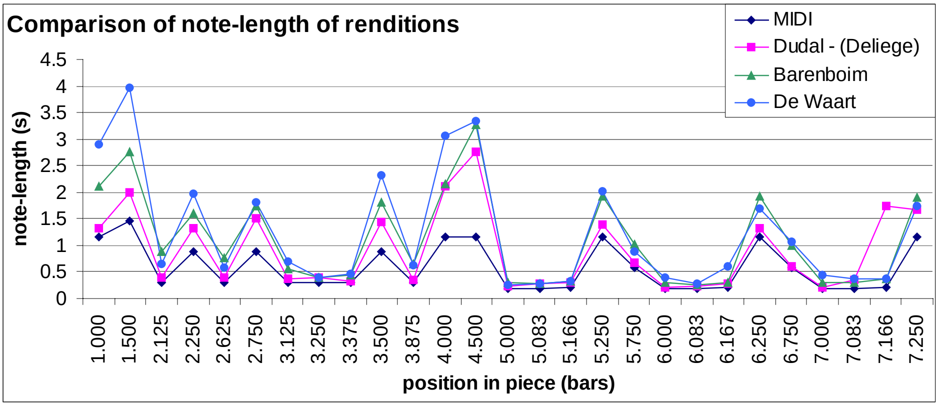
the Gestalt ideas of Même et Different (as interpreted by Deliège) are helpful in predicting segmentation points. However, it may be useful to consider a wider range of factors.

8.3.2 Deliège's method in relation to the current one

8.3.2.1 The performance

Deliège gives the time positions of the notes of bars 1-7 of Dudal's recording (p. 70), which are here compared with those used in the current study (Graph 8.3.2.1).

Graph 8.3.2.1



The overall shape and lengths of the shortest notes are the same in all the renditions. The main differences are with the longer notes. The most dramatic lengthening in the recordings is at the end of bar 3 and bar 4 (Deliège's position I). These notes may be more accentuated than in the MIDI version allowing more time for listeners to respond. The consequences are seen in the listening results (section 8.3.3).

Comparison of Deliège's results with the MIDI listening results, allows investigation of the extent to which Deliège's results, are mirrored in response to a rendition that omits performance cues. Comparison with two other performances allows an investigation of the extent of the dependence of responses on specific interpretations.

8.3.2.2 The instructions

The instructions for the subjects reveal much about what Deliège regards as segmentation (and by implication for this study, phrasing). She asks them to treat the music as a text (a story or a discourse) and to place segmentations in a manner analogous to punctuations (p. 68). This places the music squarely in the realm of language. The transfer of ideas and terminology from the linguistic to the musical mode may be reflecting Deliège's view that there are probably commonalities between visual, linguistic and musical perceptions (p. 64). Specifically, this reflects what seems to be her view of the commonality between segmentation and grouping in language and music. The linguistic terminology may also help to clarify the task to non-musicians. However, this may reduce the freedom of interpretation of what a segment of music is by confining it to a linguistic analogue.

The aims of the current study included the exploration of the listeners' definitions and identifications phrases and had a different approach keeping the tasks confined as much as possible to the musical mode and identify how the listeners understood the term (chapters 1 and 3, section 3.3.2.1 and appendices 3.2 and 3.4).

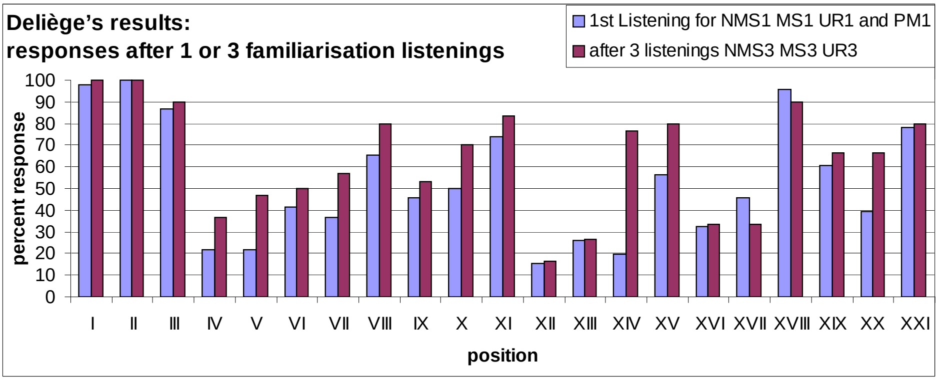
8.3.2.3 The familiarisation step

The familiarisation step allowed the subjects to get to know the piece and task before the first set of data was recorded. While listening to the piece, subjects could watch the progression of time information on a screen and note down the times at which they would mark segmentations analogous to punctuations. Two consequences of this familiarisation step are useful for data gathering but potentially affect the results.

Firstly, data is gathered after the subject has had a possibility to make decisions. The results of 'learning' or getting to know the piece, in Deliège's experiment, are observed by comparing the results of responses after three familiarisation listenings with the responses after one familiarisation listening. In Deliège's experiment there is, overall, a higher level of response for each position after three familiarisation listenings than after one (see graph 8.3.2.3). Deliège's results reveal a snapshot of responses of different subjects despite the possibility of recording their choices during the familiarisation listenings. In other words, this potentially important information on 'learning' is not revealed. In the current study, changes between listenings can be followed for individual listeners.[63]

[63] For the reconstruction of the piece from given segments (mental line task) it seems that musicians need less familiarisation listenings than non-musicians to organise the segments in their correct original order (musicians need 3 listenings, while non-musicians need 5) (p. 85).

Graph 8.3.2.3[64]



Secondly, by presenting subjects with a visual representation of the time information of the piece (presumably in a numerical form),[65] the study is not only of musical memory and learning but also of visual and numerical memory and learning. It is unclear from the paper whether or not the subjects were presented with the visual information during the task and whether or not they could refer to the notes they had made during the familiarisation task. If they did have access to both, the task becomes less one that benefits from familiarity with the piece or a memory task and more one of efficiency of decision in the first listening: the familiarisation task could become the 'experimental' run where the decisions are actually made and the second listening could become the run were the data is recorded but decisions are less likely to be made. If, on the other hand, they did not have access to the visual information or the written notes that they had made in the second listening, the task becomes one, not only of memory of or online response to the music but also one of memory of visual and numerical information. In both cases, the results do not reflect pure and immediate response to heard music.

The current study investigates the differences between the results of subsequent listenings of the same subjects and records responses from the first listening. Furthermore, in order to investigate only responses to music, visual and numerical information are not presented to the listeners (chapter 3).

8.3.2.4 Within subject consistency (and familiarisation)

By comparing the response to the simple segmentation task with the hierarchical segmentation task, Deliège hopes to show within subject consistency. This is expressed in the 'Confirmation' results (Deliège's Figure 2, p. 71). On one hand, this seems a useful and time-efficient approach – dealing with three aims (familiarisation, consistency and hierarchy) in one task. On the other hand, it seems that expecting consistency (i.e. the same subject to give the same results in

[64] Abbreviations are explained in section 8.3.1.4
[65] though this is not clear

subsequent listenings) as well as expecting to see the results of increased familiarity is contradictory.

Moreover, the simple segmentation and hierarchical segmentation may require different abilities and different approaches to listening to the music. The first can be done in a purely linear manner without reference to the past and decisions can be instantaneous, immediate and primarily independent, while the second implies the need to consider several levels of importance simultaneously, systematic differences in importance between points of segmentation, and reference to both the past and the future.

As an attempt to try the method, in a pilot part of my own study, I asked listeners, after several listenings, to attempt to provide hierarchical responses. All listeners found this task more difficult than the simple segmentation and it seemed inappropriate to compare the results of the two different tasks in a test of consistency. In investigating "hierarchy" in this study it was therefore decided to use two other approaches: to look at the overall proportion of responses and to suggest to listeners that if they heard more than one phrase structure they could represent it in different runs rather than within the same run.

8.3.3 Results of the current study in relation to Deliège's results considering the differences in methodologies

The importance of this comparison lies in that Deliège and others (including, Ferrand et al., 2002; 2003, section 8.5) base their approaches on often-quoted Gestalt principles and then base their conclusions, theories and models on these experimental results.

A comparison between the current study and Deliège's is possible because of the similar method applied to the same piece though, as described above, there are some differences in experimental set-up. The following discussion is devoted to the results of the most similar parts of the experiments.[66] The clearest direct comparison between Deliège's study and the current one is that which compares responses after one familiarisation listening for Deliège's simple segmentation task and the second listening of the current phrasing task as they are the most similar in terms of task and number of listenings.

8.3.3.1 The current MIDI results in comparison to Deliège's simple segmentations

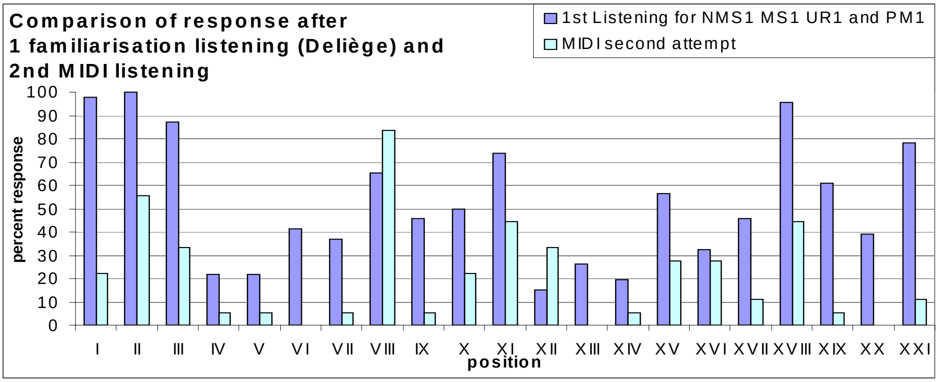
Comparison of the results of Deliège's first listening and the second listening to the MIDI rendition shows that the main response peaks coincide with only four

[66] A general discussion of the results for this piece for this study is given in chapters 3, 4, 10 and 11

exceptions: two which anticipate Deliège's positions and two pairs which merge two of Deliège's positions (chapter 10).

However, the proportion of listeners indicating phrase starts in the MIDI study is much lower than the indication of segmentation points Deliège's study (graph 8.3.3.1). Deliège's study used a recording of a real cor anglais, with the performance cues that that implies, while the MIDI version takes the note length and note pitch from the score "as written". In addition, as discussed above, the subjects in Deliège's study, unlike the listeners in the current study have had a visual representation of time and were given the opportunity to make written notes as to segmentation location.

Graph 8.3.3.1



8.3.3.2 The positions, Deliège's six and musical cues

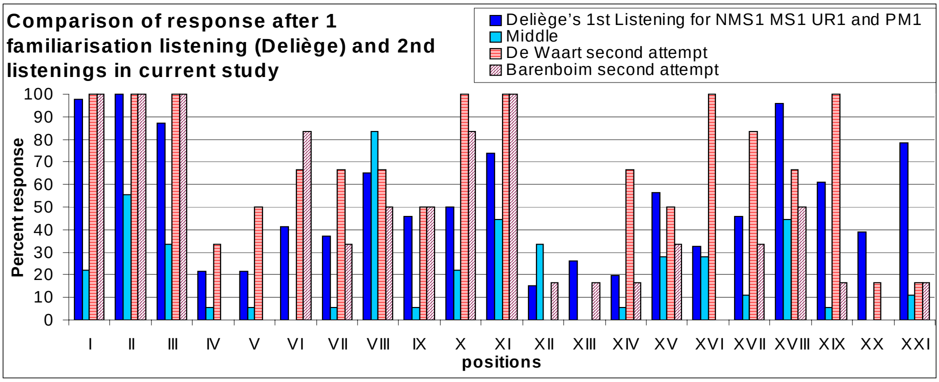
Not only is the overall response lower for the MIDI, there are less positions identified in response to the MIDI rendition than in response to Deliège's. In general, therefore, the phrases identified by listeners in the current study are longer. There are several possible reasons. For example, listeners have less time to respond because notes are not lengthened relative to the underlying pulse in the MIDI, listeners are not led in the MIDI to an interpretation favoured by the performer who may prefer shorter phrases. However, there is a further reason for this difference: listeners respond to some, possibly the clearest, musical cues when presented with the MIDI rendition (chapter 10).

All six of Deliège's predicted positions (section 8.3.1) are identified in responses to the MIDI. However, four MIDI response positions remain unaccounted for in Deliège's predictions. All those positions that have the higher MIDI responses occur exclusively at positions of musical features predicted by musical analysis carried out in this study (chapter 10)

8.3.3.3 The current performance (and MIDI) responses and Deliège's results

Comparison between responses to different performances is more direct than comparison between MIDI and performance. Graph 8.3.3.3 shows responses to the second listenings of the first sessions for the performances,[67] and the same data shown above for both my MIDI and Deliège's results. Deliège seems to consider all the responses at the positions as significant.

Graph 8.3.3.3



The responses to both of the performances in the current study have a smaller overall number of phrase positions identified than in Deliège's. The number of positions is higher in response to De Waart than in response to Barenboim, which has the same number of positions as to the MIDI. These differences may be because the responses are to different performances emphasising different aspects of the musical structure. However, the majority of positions responded to in the MIDI are included in responses to the performances. This systematic commonality can be explained by musical cues (chapter 10).

There are several possible reasons for the difference in response between the studies: 1) The tasks were described differently, for example, in Deliège's experiment, the task was described as a punctuation one, while here it was described as a phrasing one. 2) The different cues provided by the different performances. Graph 8.3.2.1 (using the information provided in Deliège's paper) shows note-length information and for first segmentation position only. This is insufficient for assessment of the reasons for the pattern of responses. Analysis of the performance used by Deliège might shed more light on the differences in phrase perception.[68] It seems that further discussion of the differences between the results obtained in the current study and those obtained by Deliège is precluded as the reasons for the choice of these 21 positions are not explained.

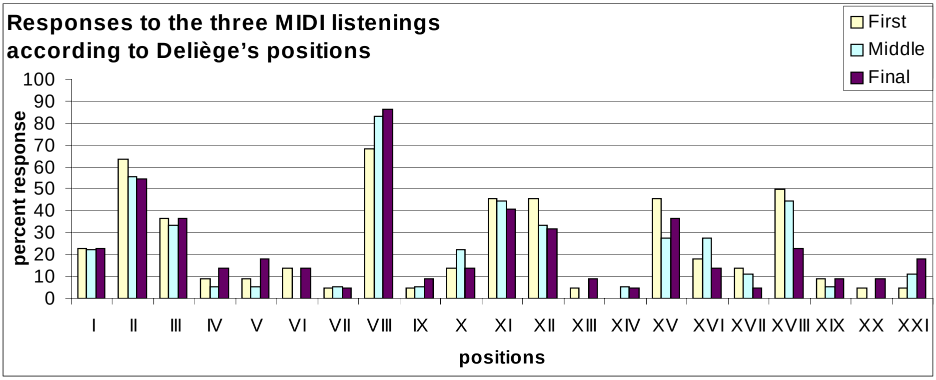
---

[67] From the session I; listeners have not heard the other performance (chapters 3 and 4)
[68] However, this has not been possible as I was not granted access.

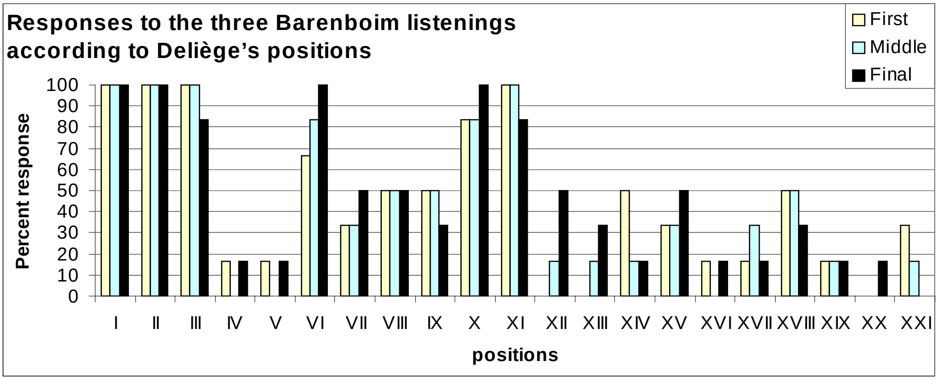
8.3.3.4 Familiarisation step

In Deliège's experiment, overall, there is a higher level of response after three familiarisation listenings than after one (graph 8.3.2.3). However, the current MIDI and Barenboim results from the three listenings indicate that there is no clear increase in responses over the listenings or even from the first to the third (graphs 8.3.3.4.1 and 8.3.3.4.2). For the De Waart there are several positions that show increase in response across listenings (graph 8.3.3.4.3). However, overall these results do not show a consistent increase in response between listenings.

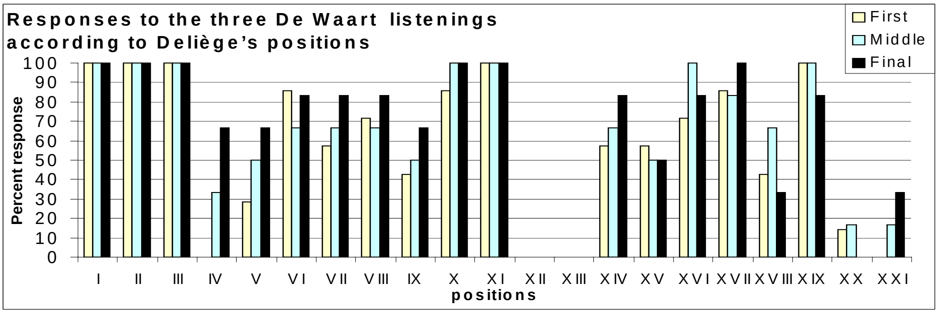
Graph 8.3.3.4.1



Graph 8.3.3.4.2



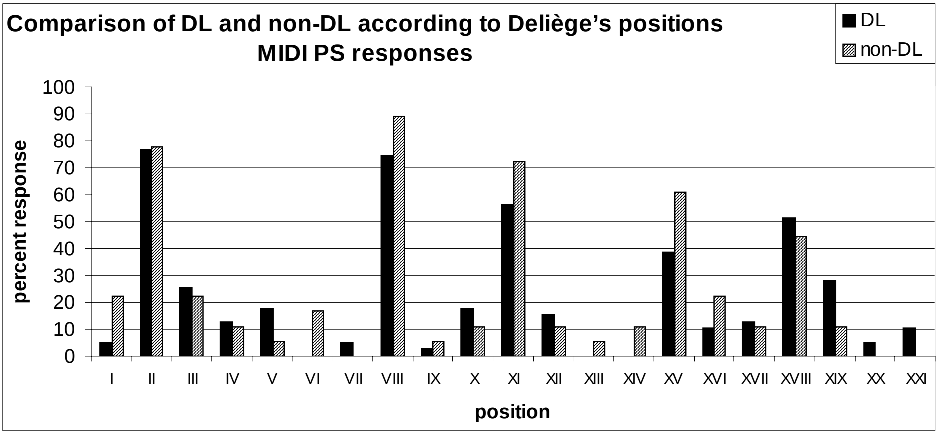
Graph 8.3.3.4.3



8.3.3.5 Musical Education

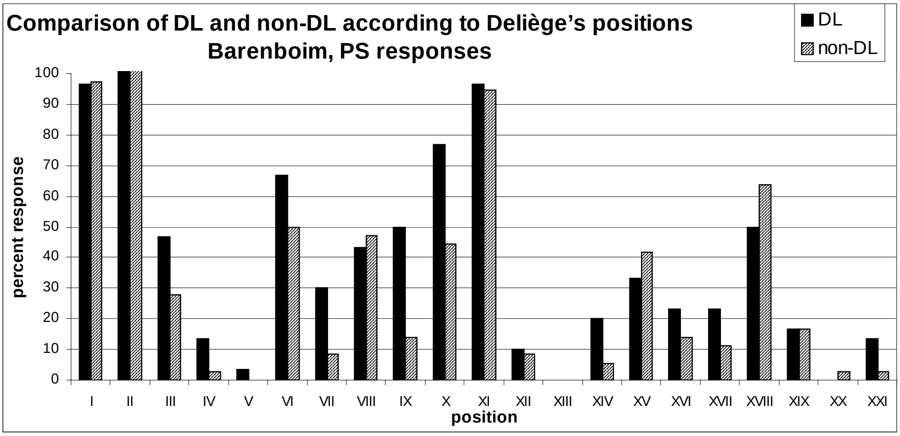
Deliège's study shows that decisions taken by professional musicians are more stable regardless of number of familiarisation listenings (p. 82).[69] However, for the current MIDI, De Waart and Barenboim PS responses for these positions, there are no significant differences between musicians (DL) and non-musicians (non-DL) ($p \geq 0.05$) (graphs 8.3.3.5.1-3).

Graph 8.3.3.5.1

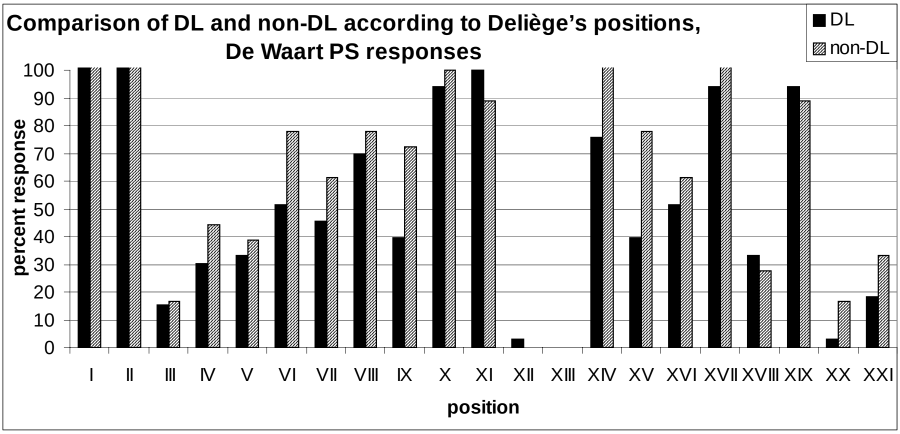


[69] For non-musicians, the secondary segmentation points are less consistent than the primary ones (p. 82).

Graph 8.3.3.5.2



Graph 8.3.3.5.3



Indeed, for some positions, particularly in the MIDI and De Waart, the response is higher among the non-DL (non-musicians) than the DL (musicians).

8.3.4 Summary of Deliège's results and findings relevant to the current work

Deliège's study was investigated in detail for several reasons. It tackles the question of segmentation from three perspectives: through simple or linear representation, hierarchical representation and memory (mental line). The study concentrates on one Gestalt principle, uses a similar methodology to one of those used in the current study, uses one of the pieces investigated in my study and concentrates on the effects of amount of musical education of the subjects and their familiarisation with the piece. In more general terms, Deliège's study seems to be both a point of reference for her own and others' theories and models, and refers to frequently-quoted Gestalt principles.

In terms of musical education, Deliège's study shows that decisions taken by professional musicians are more stable regardless of number of familiarisation listenings (p. 82). The same systematic trend does not seem to occur in the current study. In terms of familiarisation for non-musicians, the more they have heard the piece, the more they segment while for professional musicians, the more they have heard the piece, the more important secondary segmentations become (p. 82). Looking at the general results (chapter 3), this again does not seem to be supported in the current study.

Overall, however, Deliège reports that the musical education and degree of familiarisation with the piece seem to have a marginal effect on the results of the segmentation tasks (1998, p. 83). This implies that subjects from different musical backgrounds segment in a similar way and are similarly consistent. In other words, the ability of segmentation is common to all subjects independent of musical training. This seems to be for the most part supported in this study (chapters 3 and 10).

Of the 21 positions discussed in Deliège's study, 10 have high responses. The current MIDI responses coincide with all 10. The responses to the De Waart performance coincide with all 10 and add 4 positions. The responses to Barenboim, while being the same number as in response to the MIDI, omit two positions high response positions and have two additional ones. Furthermore, Deliège states that she can provide explanations for only 6 of these using the Même-Different principles. The MIDI and other responses here confirm the predictions made on the basis of these Même-Different principles. The other positions of response will be considered in chapter 10.

8.4 Cambouropoulos' boundary detection and segmentation

8.4.1 Introduction

Cambouropoulos developed two models that describe different aspects of segmentation. One approaches the question from the perspective of identifying musical cues specifically related to the boundary (Local Boundary Detection Model, henceforth, *LBDM*), while the other approaches the question from the perspective of the content of the segment, particularly repetition (Pattern Boundary Detection model). Here, the two models are discussed separately and then compared.

8.4.2 Boundary Detection

8.4.2.1 Cambouropoulos' Local Boundary Detection Model

Aim / Structure

The *LBDM* calculates boundary strength for each interval of a melodic surface according to the strength of local discontinuities; peaks in the resulting sequence of boundary strengths are taken to be potential local boundaries.

The *LBDM,* according to Cambouropoulos (2001), is simpler and more general than the models of Lerdahl and Jackendoff (1983) and Tenney and Polanski (1980) (chapter 1), both of which he sees as special cases of the *LBDM*. The *LBDM* accounts for '*any* change in interval magnitudes' (Cambouropoulos, 2001, p. 2)[70] and is based on two rules.

> '**Change Rule:** Boundary strengths proportional to the degree of change between two consecutive intervals are introduced on either of the two intervals (if both intervals are identical no boundary is suggested)' (2001, p. 2). In principle the change can be in pitch, time, dynamics, harmony or any other parameter relevant for the description of melodies. However, only pitch, Inter-onset Interval (IOI) and rest information are used in this study' (2001, p. 2).[71]

[70] In this section, Cambouropoulos' papers are referred to with date and page number.
[71] Inter-onset interval refers to the length between the beginnings of two consecutive notes. Offset-to-onset intervals refer to the length between the end of one note and the start of the next.

> '**Proximity Rule:** If two consecutive intervals are different, the boundary introduced on the larger interval is proportionally stronger' (2001, p. 2).

Cambouropoulos explains that the 'current implementation of *LBDM* is not expected to find all local boundaries correctly as it does not include harmonic profiles of melodies and also it does not take into account melodic similarity which is paramount for establishing important groups of notes' (2001, p. 2).

The groups that the boundaries separate

The *LBDM* not only enables the detection of local boundaries in a melodic surface, but can be used for musical segmentation (2001, p. 1). Cambouropoulos hints both as to the approximate lengths of the sections between the local boundaries, and the musical name of these sections. He distinguishes between weak and strong boundaries and explains that strong boundaries constitute roughly 25% of all notes, i.e. roughly 1 in 4 (2001, p. 4). This is supported in the graph showing results (Figure 2, p. 7). Though this is somewhat contradicted by his quotation of Friberg et al., (1998, p. 272) which states that smaller units, typically consist of 1 – 7 notes (2001, p. 4). In terms of the musical category, Cambouropoulos implies that there is a distinction between a melodic phrase (higher level) and the sections that he is investigating ('lowest level'): 'It is commonly hypothesised that the ending of a musical group, such as a melodic phrase, is marked by a slowing down of tempo, i.e. relative lengthening of the last notes (Todd, 1985). For musical groups at the lowest level, i.e. small melodic gestures of just a few notes, it is commonly assumed that the final note IOI is lengthened and a small micropause inserted' (Friberg et al. 1998)' (2001, p. 1).

8.4.2.2 Cambouropoulos' Computational and Experimental Method and Results

Cambouropoulos (2001) reports the results of a number of computational and performance-based experiments. The only one with a musical example in the paper itself is the experiment on 22 performances of the first 20 bars of Chopin's Etude Op. 10, No. 3 and is therefore the one that can be most directly related to the current study. According to Cambouropoulos, this piece has a rather clear low-level grouping structure, which is determined by rather long notes in between shorter ones (2001, p. 5). The performers presumably played all the parts of the score but Cambouropoulos computes an IOI deviation curve for the melodic part only.

Cambouropoulos states that there are 17 boundaries 'indicated by relatively long notes in between short ones' (2001, p. 5) for a total of approximately 98 notes of the melody. The *LBDM* model detects correctly all the 'important' boundaries with one exception (between bars 15 and 16). Cambouropoulos studies the 11 strongest of these. Second-to-last note lengthening (delay of the last note) in the performances coincides with the majority of these. Moreover, notes with stronger *LBDM* values tend to be lengthened (2001, pp. 5 – 6).

8.4.2.3 Discussion of Cambouropoulos' *LBDM* aims, methods and results relevant to the current work

According to Cambouropoulos, '[i]t is clear that the *LBDM* is not a complete model of grouping in itself, extensions of the current model (e.g. harmonic component) and also complementary models for establishing musical groups via melodic similarity are necessary' (2001, p. 4). Bearing this in mind, however, this paper is relevant to the current study for a number of reasons:

Cambouropoulos has a clear aim of identifying boundaries of groups. Most theorists dealing with grouping emphasise the importance of the identification of groups but in practice concentrate on boundaries only. However, the 'level' that Cambouropoulos is looking at is limited to sections of approximately 4 notes, or maybe 1–7 notes, referred to as musical gestures which are shorter than phrases. This is supported by the results of the current study, which showed that most phrases identified by listeners are longer than 4 notes (chapters 3, 4, 6 and 10).

Cambouropoulos is using two common Gestalt principles: Proximity and Change and looks for their expression as musical cues. These are often used in other studies (see the other studies in this chapter) regardless of the size of the group being considered. Having used these cues in the context of these Gestalt principles and having done so on a variable scale (all distances and relative distances are considered), Cambouropoulos shows that they can be successful for boundary detection of musical gestures.

It seems from Cambouropoulos' description of the model that once a feature is identified, a boundary is placed and this cannot be overruled and removed by other information. Instead there is only an expression of relative strength of the boundary. The strength of the boundary is determined by the strength of the feature. However, the strength of the feature is determined by its strength relative to other examples of the same feature. One implication is that a spread of more features results in smaller gesture lengths. The tendency for smaller and smaller groups seems not to be problematic for Cambouropoulos. This is seen by the first way in which he 'improves' the results of the *LBDM*: by allowing 'groups of *one* note' (2001, p. 3). This contradicts Lerdahl and Jackendoff's approach to grouping in which they prefer larger groups (GPR 1, Lerdahl and Jackendoff, 1983, p. 345, chapter 8.1).

Cambouropoulos suggests that boundaries may be of different strengths. It could be, as is often suggested in hierarchical approaches (such as Lerdahl and Jackendoff, 1983), that the stronger the boundary, the more important it is at the next hierarchical level. Concurrently, Cambouropoulos reminds us of the theory (as described for example by Todd, 1985) that longer IOIs in performance reflect stronger boundaries. Therefore, if stronger boundaries as predicted by the *LBDM* coincide with stronger boundaries as suggested by IOIs (as they do in his study), we may be able to use these cues to identify boundaries at different hierarchical grouping levels and possibly at the phrase level.

The comparison of the experimental study of 10 performers' IOIs in the first 10 beats of Chopin's Etude Op. 10 No. 3, with the results of the *LBDM* on the same section of music, however, shows that the relationship between the two is complex. As can be seen from the graph of the first 10 beats of the piece (replicated below as table 8.4.2.3.1), the *LBDM* identifies 6 boundaries: the two weakest ones are on beat 5 and after beat 6, the three strong ones are on beats 1, 3 and 7 and a weaker one on beat 9. In other words, if arranged from weakest to strongest the boundaries are: after beat 6, on beat 5, on beat 9 and on beats 1, 3 and 7 (table 8.4.2.3.1). At the same time, the results for the 10 performers show a second-to-last note lengthening before each one of these (from the shortest IOI to the longest IOI): before beat 1, before beat 5, beat 6, before beat 7, before beat 3 and before beat 9. In terms of boundary strength, this implies the following order:

Table 8.4.2.3.1 showing beat number of boundaries in order of strength from weakest to strongest according to *LBDM* and performers' IOIs

Boundaries from weakest to strongest according to LBDM	***Performers' IOIs from shortest dips to longest***
6	1
7	5
9	6
1	7
3	3
7	9

Although the same boundaries are identified by the two methods, the relative strength according to the *LBDM* and the performers' IOIs coincide only at bar 3. Moreover, the longest IOI for the performers does not coincide with the most important *LBDM* boundary (beat 9 is the ranked third out of the 7).

In terms of the *LBDM* model, the boundary at beat 9 is weaker than those at beats 1, 3 and 7. This is presumably because the long note starting on beat 9 is shorter by a semiquaver than those starting on beats 1, 3 and 7, resulting in a smaller note-length difference for beat 9, than for the others.

There are other musical reasons that may support the stronger boundary on beat 7 than on beat 9,supporting the *LBDM*. For example, from the perspective of a listener who does not know the piece, 7 is in the fourth bar, while 9 is in the fifth and, if the idea of 4–bar sections are stronger than 5 bar sections, the section should end at 7. In this case the section would end on the dominant.

In retrospect though, the section between 7 and 9 could be heard as a close of the opening passage (returning to the tonic), in which case the stronger boundary is on beat 9. Moreover, the comparison here is not with listeners but with performers who (presumably know the piece) and wait for 9 for the lowest IOI. Only at 9 is harmonic resolution achieved, the theme has ended (and the pitch gap occurs between 9 and 10 and not between 7 and 8) leading to a stronger boundary on 9 than anywhere earlier in the piece.

It seems, therefore, that, although the *LBDM* is promising due to its use of musical cues and its range of boundary strengths, it may not be directly applicable to the investigation of phrasing. For this purpose, and even for the detection of boundaries, as Cambouropoulos says, more cues may be needed. Moreover, even for phrase *boundary* identification, the relationships between consecutive phrase boundaries and the internal structure of the phrases may be useful as indicated by the relationship between beats 7 and 9 above.

8.4.3 Segmentation

8.4.3.1 Cambouropoulos' Similarity Model (PAT)

Aims

In Cambouropoulos (2003) 'melodic pattern extraction is used as a means to segment a melodic surface' (p. 134). This is done by identifying 'significant' repeating musical patterns (2003, p. 134), i.e. searching for similar patterns in the same piece.[72]

The focus of this study is 'primarily a special case of melodic similarity, namely immediate repetition of melodic passages' (2003, p. 135). Such repeating passages are often characterised by: 1) divergence towards their endings, 2) small variations, and 3) transposition of the repeated passage (2003, p. 135). There are several versions of the model (elaborated below) one of which only takes account of the beginning of the repeated segments, accounting for characteristic (1). This also allows characteristic (2) to be accounted for as long as the 'small variations' are not at the start of the segment. Characteristic (3) is automatically accounted for as the pitch is taken as interval information (2003, p. 135).

Cambouropoulos sees musical similarity as a higher-level process than local discontinuities searched for by the *LBDM*. He explains that similar musical patterns tend to be highlighted and perceived as units/wholes whose beginning and ending points influence the segmentation of a musical surface (2003, p. 135).

[72] See also Cambouropoulos (1998a, 1998b and 2006).

The cues

Cambouropoulos' model looks for repetitions of pitch intervals and IOIs (2003). For two of his examples (Figure 1. *Frère Jacques* and Figure 2. *Choral St. Antoni* (Brahms, *op. 56*), p. 137), of the two cues, pitch intervals are more appropriate for segmentation because the rhythmic profile is poor in terms of information content; the number of available duration values is too small and repetitions too many (2003, p. 135). For his final example (Figure 3. *Theme of section III of Mozart's G minor Symphony*, p. 137), Cambouropoulos suggests that 'a more abstract representation for pitch intervals may be useful, such as a step-leap profile, especially if it is coupled with duration information' (2003, p. 136). Cambouropoulos (2006) shows that results can be further improved for some pieces if the abstract symbolic representations of pitch intervals can become more flexible in terms of category gradedness and membership. For example, thirds can be either steps or leaps. A further development is that almost all other intervals can belong to more than one sub-group, including, small, medium and large leaps (2006, pp. 15-19).[73]

The 'level' of musical structure

Cambouropoulos assumes that similarity for melodic segmentation is limited to the melodic surface because 'extracting patterns at reduced versions of the melodic surface would result in ambiguous segmentations, as it would not be possible to define where exactly the boundaries of the repeated patterns should be placed (since there are notes missing from the reduced version)…Of course, musical similarity appears in many guises at deeper levels of musical structure but in such cases it is likely that this sort of abstract similarity is not the most crucial factor in segmentation tasks – other factors such as gestalt-based local boundary detection factors or learned schemata (e.g. harmonic cadences) are responsible for segmenting the surface and only then are more sophisticated comparisons of segments made possible at more abstract levels of description' (2003, p. 135).

Significant repeating musical patterns

For Cambouropoulos, 'significant' 'patterns are defined primarily in terms of frequency of occurrence and length of pattern' and patterns should not overlap (2003, p. 134). This is expressed in the following rules: 'a) prefer most frequently occurring patterns, b) prefer longer patterns and c) avoid overlapping' (2003, p. 134). The combination of these rules can be expressed as one figure obtained from a 'Selection Function' (2003, p. 134).

[73] Cambouropoulos (2006) compares the results of this version of the model with an empirical study (Koniari et al, 2001) and shows that the model identifies the same segments as the listeners.

In Cambouropoulos' model all discovered patterns merge into a single 'pattern segmentation profile' that signifies points in the musical surface that are most likely to be perceived as points of segmentation. This is done through a pattern boundary strength profile (*PAT*), which uses the Selection Function and identifies all the patterns that begin or end at each position (in a modification discussed below, it looks only at the patterns that begin at each position). No pattern is disregarded. Instead, all contribute to a possible boundary by a value proportional to the score obtained by the Selection Function. The hypothesis is that points of local maxima are more likely to be perceived as boundaries because of musical similarity (2003, pp. 134-135).

The resulting segments and boundaries

In Cambouropoulos (2003), the relation between segmentation and phrasing is hinted at through a reference to David Lidov (1979) who calls the kind of repetition that Cambouropoulos is dealing with as '*formative repetition*', whose 'function is to establish or to form motives and phrases' (2003, p. 135). However, it seems that the segments Cambouropoulos is mainly referring to are smaller than what, according to, for example, the MIDI responses of this study (chapter 3), can be regarded as a phrase.

Cambouropoulos describes the module of Cambouropoulos (1998a and 1998b, the predecessor of the current *PAT* model), as complementary to the *LBDM* (2001, p. 4). It may be that the *LBDM* and the *PAT* are complementary too. However, there is a difference in aim: The *LBDM* seems to concentrate on the boundaries whilst the starting point of the *PAT* model is to regard the similar musical patterns as 'units/wholes whose beginnings and endings influence the segmentation of the musical surface' (2003, p. 135), implying that the beginnings and endings are secondary. In a modification to the model, 'only the beginnings of patterns contribute to the strength of the pattern boundary profile' (2003, p. 136) and by implication, the segmentation and Cambouropoulos calls the later 2003 model a 'pattern boundary detection model' (2003, p. 136). In addition, in practice the boundaries identified sometimes coincide, and sometimes do not. The *LBDM* and the *PAT* may be complementary and the boundaries identified by the two models may be complementary too (and so equally valid). It would be interesting to explore the relative strength of the boundaries identified by the different models.

8.4.3.2 Discussion of Cambouropoulos' Similarity Model (PAT) aims, methods and results relevant to the current study

Cambouropoulos' work has been discussed here as it is one of the few that attempts to formalise the very difficult feature of repetition (others include, Steedman 1977). As with the *LBDM* model (section 8.4.2.4) it may be possible to use these cues to identify boundaries at different levels and possibly at the phrase level.

In the following discussion it should be noted that Cambouropoulos superimposes results of the *LBDM* model and the different versions of the similarity (*PAT*) models. However, just like in the *LBDM* and the *PAT* models, the relative importance of final features (step-leap interval vs. duration ratio) is not explicitly given. Cambouropoulos explains that '[t]he total boundary strength profile can be calculated as a weighted average of the local boundary and pattern boundary strength profiles even though more sophisticated methods for combining the two should be explored' (2006, p. 21). So, for the following discussion, it is assumed that if the two models are combined, and they find different segmentation positions, the positions from both models are as important as they would be independently. The implication from this is that, for now, potentially, the number of segmentation points (however weak) increases.

The boundary identified in Figure 3. *Theme of section III of Mozart's G minor Symphony* (2003, p. 137) by the step-leap and duration ratio profile is the clearest among the three examples given and coincides with what could be a phrase boundary, or at least a subphrase.[74] At this position, in addition to the beginning of the transposed three-bar repetition, there is an arrival on the dominant note (and presumably dominant harmony) in the form of an imperfect cadence and the opening note is returned to.

Similarly, using the model that looks only for beginnings of patterns, the peak identified for Figure 2. *Choral St. Antoni* could be a phrase start. This example, however, shows the difficulty of concentrating on boundaries, and if looking only at boundaries, the difficulty of limiting the search to either beginnings or endings: The previous phrase 'ends' on the c, one beat earlier, and then there is an upbeat, or "filling", of 3 semiquavers. It is possible that, as in the Mozart Sonata used in the current study (chapters 3 and 10) the musical 'solution' is not limited to one note; listeners and performers may identify different possible boundaries at different positions in this area depending on the cues they respond to. Even taking this into account though, the success of the model here, not only for its original intention in searching for similar units and segment boundaries, but also for application in the search for phrase boundaries is promising. For both of these examples, the *LBDM* does not perform as well as the *PAT* model in a search for 'phrase' boundaries.

For the first example, Figure 1. *Frère Jacques*, the repeated segments are much shorter (3–6 notes or one bar) than in the other two examples, so it is not surprising that the segments identified are shorter. However, the relative importance of the different boundaries as presented by both the *LBDM* and the *PAT* model (the latter model looks here only at the pitch interval profile for both beginnings and endings) is different from what might be expected. For example, for the *PAT* model the boundary between what might be expected to be the

---

[74] Though experiment results are not available for these pieces.

longest groups (bars 1 – 4 and 5 – 8) does not have the highest figure in the boundary strength profile (there are three stronger boundaries). Conversely, for the *LBDM* this position is the highest. However, other positions (bars 4 and 8) are more important than the preceding bars (3 and 7 respectively) thus differing from the ones that one might expect to be the stronger boundaries. This is because in *Frère Jacques* there are many cases of immediate exact repetition - the type of repetition that the model is designed to identify - but here they are so common that once the exact repetitions are identified, at least on some level, they are grouped together and distinguished from the different sections that follow. In other words, having grouped according to similarity on one level, thereafter, difference rather than the similarity between the pairs indicates segmentation. The importance of difference is often quoted in the representation of Gestalt principles in musical terms (for example, Deliège, 1998, section 8.3). Cambouropoulos has begun to tackle the idea of 'change' (2001, p. 2) though in a slightly different way; changes from one note to the next rather than between groups.

This is not to say that repetition is not an important musical cue for segmentation. It suggests that immediate repetition, in terms of phrasing at least, is useful only if the segments repeated are relatively long and that short segments, when repeated, become grouped together within larger sections (phrases) that contrast with other material. All the repetitions identified in the pieces used in the current listening studies coincided with the majority of listener responses and repetition seems to be one of the most important musical features (chapters 10-14).

The above discussion suggests that any one of the cues studied in these two models (change, proximity, similarity) or maybe other cues (including difference in melodic material) may be important under different circumstances. One of the main challenges is the identification of the circumstances i.e. cue combinations, for which different cues are most important for segmentation, or more specifically in this study, phrase identification.

8.5 Bod’s Memory-Based models of melodic analysis

8.5.1 Introduction
8.5.2 Method: Training and testing using the Essen folk-song collection
8.5.3 Results of the memory-based approach
8.5.4 Discussion of aspects of Bod’s DOP theory and results relevant to the current work

8.5.1 Introduction

Bod (2001; 2002), like Temperley (2001, discussed in chapter 8.7), presents an approach to melodic phrase structure that aims to identify phrases in monophonic folk songs. Unlike Temperley, Bod argues for a completely ‘memory-based approach to music analysis which works with concrete musical experiences rather than with abstract rules or principles’ (Bod, 2002, p. 27).[75] For him, analysis seems to imply the ability to computationally identify musical phrases already identified by a musically experienced annotator. He therefore develops a memory-based probabilistic program that learns the series of notes that make up phrases identified by the annotators along with their frequency of occurrence. The program, when applied to un-annotated pieces identifies phrases on the basis of what has been “learnt.” The pieces used are folksongs from the Essen folksong collection; monophonic pieces for which the text is not usually given but for which relative pitch, note length and annotator’s phrase marks are. The annotation of phrases, and therefore the learning and subsequently identified phrases in the test pieces of the same genre, are annotated with a single level of phrasing. Moreover, the method is based on learning series of notes and hence, melodic, taking into account pitch relative to the tonic and rhythm. This is similar to the methods used in other studies, such as Temperley (2001, see chapter 8.7).

According to Bod, the results of his program indicate that there are ‘grouping phenomena that challenge the commonly accepted Gestalt principles of proximity, similarity and parallelism’ and that these ‘grouping phenomena can neither be explained by other musical factors, such as meter and harmony’ (p. 27). The ‘results showed that there is a class of musical patterns, so-called jump phrases, that challenge both the Gestalt principles of proximity and similarity and the principle of melodic parallelism. Jump-phrases provide evidence that grouping boundaries can appear *after* or *before* large pitch intervals, rather than *at* such intervals, and that grouping boundaries can even appear between *identical* notes (that are preceded and followed by relatively large intervals). We have seen that Gestalt-based, parallelism-based and/or harmony-based models are inadequate to deal with these gradient phenomena of music analysis since they can capture the entire continuum between jump-phrases and non-jump-phrases’ (Bod, 2002). This implies that memory of previous experience of pieces is so strong that we learn

[75] In this chapter, this text is referred to with page numbers only.

the note combinations associated with phrases. Similar difficulties with application of Gestalt principles, particularly by listeners with less musical experience, have also been identified with pieces from the same corpus (Schaefer et al., 2004). Bod uses the terms phrase and group interchangeably and groups seem to be defined by the annotators' markings (p. 28).

8.5.2 Method: Training and testing using the Essen folk-song collection

In Bod's approach new 'pieces of music are analyzed by combining fragments from structures of previously encountered pieces' (p. 27). This is done using the 'occurrence-frequencies of the fragments … to determine the preferred analysis of a piece' (p. 27). Unlike all the other approaches discussed here, and in a way similar to approaches used in natural language processing (such as Manning and Schütze, 1999), this is a 'supervised, memory-based approach to music analysis which works with concrete musical fragments rather than with abstract formalisations of intervallic distances, parallelism, meter, harmony or other musical phenomena'. This picks up on recent psychological investigations that suggest that more frequently encountered fragments are better represented in memory (Saffran et al., 2000) and therefore more easily activated than less frequently encountered ones (pp. 27-8).

Bod uses the Essen folksong collection both for training (5,251 pieces) and testing (1,000 pieces) while pointing out some of the difficulties with the annotation of the phrases in the collection (p. 29). For example, Bod gives an example of a folksong for which each phrase could have subphrases (p. 29). Bod tested three memory-based parsing models using different techniques: Treebank grammar (Charniak, 1996), Markov grammar (Collins, 1999; Seneff, 1992) and Data-Oriented Parsing (DOP) (Bod, 1993; 1998). The DOP technique (combined with a Markov technique) records all cases of annotated phrases as rewrite rules and the probabilities of finding them in the different contexts in the different folksongs, using a history of four notes (p. 31). The 'most probable parse of a folksong is … computed by maximising the product of the rule probabilities that generate the folksong' (p. 31). The results of using this DOP-Markov parser are discussed below.

8.5.3 Results of the memory-based approach

Bod tested his approach against a set of 1,000 manually annotated folksongs from the Essen folk-song collection. His DOP-Markov parser 'obtained a precision [correct identification out of annotations] of 76.6%, a recall of [proportion of correct out of proportion identified] 85.9% and an F-score [combination of the two] of 81.0%' (p. 29). Moreover, as described above, Bod discusses 'jump phrases' (phrases that begin or end with a leap, usually from the dominant to the tonic) and explains that other methods (following Gestalt principles, metrical or harmonic features) would not identify these phrases correctly.

Bod also investigates the minimum size of training set necessary for good test results. He shows that there is a rapid increase of success up to 2000 folksongs and then a slower increase up to 5251, indicating that a minimum exposure of some size to "phrased" music is necessary for phrases (in the same genre) to be identified. This method is, at best, simulating adult perception (p. 34). Addressing the question of how this knowledge is acquired in the first place, Bod suggests that this may be through a bootstrapping process in which the 'discovery of similar recurrent patterns and distributional regularities plays an important role. As soon as a pattern appears more than once, it may be hypothesised as a group, and may be used as a productive unit to analyse new pieces' (p. 34). Furthermore, Bod explains that the frequency with which a pattern occurs is used to decide between conflicting groups.

8.5.4 Discussion of aspects of Bod's DOP theory and results relevant to the current work

This approach is specifically for monophonic folksongs and for a case in which there is one parse per song. As there can be more than one annotation for each one of the pieces used in the current study (as shown by the results presented in chapters 3-4), it is difficult to directly apply this method to the pieces of the current study. Therefore, the present discussion concentrates on the theoretical considerations of Bod's approach.

In Bod's approach, ambiguity is considered to be limited to the "computer's" analysis and each listener is assumed to have only one interpretation (p. 27). Despite the success described by Bod, he does suggest that a purely memory-based model may not suffice (p. 28). However, leaving aside the questions of uniqueness of the phrase identifications in the Essen folksong collection, particularly without considerations of the texts or musical norms of upbeats, the results indicate that the other approaches (Gestalt, harmonic, metric) to phrase identification do not identify all phrases identified by the annotators. Moreover, Bod's method is different from the other studies in that it takes the phrase as a whole and not just the boundaries or starts and ends indicating the importance of the internal patterns of the phrase.

8.6 Ferrand's unsupervised learning of melodic segmentation

8.6.1 Ferrand et al.'s probabilistic model of melodic segmentation
8.6.2 Ferrand et al.'s segments
8.6.3 Comparison between Ferrand et al.'s experimental and computational results
8.6.4 Discussion of aspects of Ferrand et al.'s theories methods and results relevant to the current work

8.6.1 Ferrand et al.'s probabilistic model of melodic segmentation

Aim and musical cues

Ferrand et al. (2002) aim to develop a probabilistic model of melodic segmentation as perceived by a listener. They do so on the basis of four musical cues: pitch step (interval distance between consecutive pitches), pitch contour (direction of the pitch step, up or down between consecutive pitches), duration ratio (between consecutive events), and duration contour (relative length) (p. 4). Though they suggest (on the basis of Krumhansl's work) that there is no evidence to distinguish quantitatively between the contributions of the different musical cues (2002, p. 2), in Ferrand et al. (2003), pitch and duration contour are excluded.

Probability, Entropy change, Predictability and Expectation

Ferrand et al. associate feature salience with expectation. They use entropy to measure unpredictability associated with different musical cues. 'Low entropy usually means high predictability but if a particular feature (e.g. note duration) is highly predictable throughout the piece then it may be because it is either highly invariant or because it follows a monotonous variation pattern.' Musical parameters with varying entropy during the piece are more informative than those with consistently high or low values (2002, pp. 2-3). Therefore, they analyse the entropy changes along a piece; transitions between high and low entropy constituting salient moments in a listening experience.

High predictability is associated with a low level of entropy change and the expectation is continuation of the same and therefore no expectation of segmentation. They aim to use these musical cues in Markov and entropy models to predict notes on which listeners in experiments have identified segment boundaries (2002).

8.6.2 Ferrand et al.'s segments

Ferrand et al. remind us that segmentation has been seen as an important part of music understanding. Moreover, they make the connection between segments and motives, phrases or sections, implying that phrases are a type of segmentation (calling them 'constituent units' amongst motives and sections) but do not elaborate on this (2002, p. 1).

It seems, however, that a segment, for them, does not have a clear identity or definition. A hint of its identity is from the results of listener experiments. However, even this is problematic. In Ferrand et al. (2003) the pieces used in the listener experiments are a selection of pieces from the Essen Folk Song Collection and two excerpts from Mozart Piano Sonatas (heard in MIDI in the experiments), and a performed version of Debussy's Syrinx. In the experiments, the methodology used is similar to Deliège's simple segmentation task with two familiarisation listenings and then two data recording runs (1998). In Ferrand et al. (2002), the experimental results are those provided by Deliège on *Die alte Weise* (1998, chapters 8.3 and 10).

According to Ferrand et al., (2002), Deliège's experiment revealed 8 main segment boundaries (identified by most subjects) and an additional set of 13 weaker boundaries. Ferrand et al. use only the stronger boundaries but with no explanation. From their graphs (2002, p. 6) it seems that they are referring to Deliège's segmentation points I, II, III, VIII (or possibly IX), XI, XV and XVIII, while for Deliège, the main segmentation points used in her third task (Deliège, 1998, p. 79) are I, II, III, IX, XI, XVIII, and XXI (and her six predictable positions are I, II, III, IX, XI, XVIII). The difference being that Ferrand excludes position XXI (the end of the piece anyway) and includes position XV (see appendix 8.3 and chapter 8.3). This comparison shows the difficulty encountered in deciding which the most important segment boundaries are, even when the piece, performance and results under discussion are exactly the same, and even parts of the contributing cues are considered to be the same (both talk of Gestalt principles). As Ferrand et al. (2002) use the same piece as used in the current study and as this is discussed elsewhere (chapters 8.3 and 10), this discussion of the experiments concentrates on the comparison between the experimental and computational results.

8.6.3 Comparison between Ferrand et al.'s experimental and computational results

Ferrand et al. explain in their conclusions that there is a close coincidence between the entropy change areas for the different cues (whether four or two, Ferrand et al. 2002 or 2003 respectively). In Ferrand et al., 2002 this is especially the case for temporal cues and in Ferrand et al., 2003 this is true of both cues used, suggesting that a change in pitch step, pitch contour (direction), duration ratio or duration contour (direction of change in direction) can indicate segmentation points as perceived by listeners.

The graph of the experimental and computational results from the Syrinx (as well as the breath marks) shows that indeed from the 14 boundaries identified in the experiment, '11 were predicted correctly by the model'; Five from the pitch information and six from the duration information (Ferrand et al., 2003, p. 4). The model also generated 5 boundaries not chosen by listeners, 3 from pitch information and 2 from duration information. It seems that pitch and duration information do go some substantial way to predicting segment boundaries identified by listeners. However, there is some over-prediction and some boundaries chosen by listeners are missed by the model.

In the case of *Die alte Weise* (graphs in Ferrand et al., 2002, p. 6), the pitch contour entropy seems least variable (and therefore least useful). In the rest of the graphs, some of the boundaries coincide with changes in entropy, some changes in entropy occur before the boundaries identified by listeners (such as in position XV of the duration ratio entropy) and some changes in entropy follow boundaries (such as in VIII of the duration ratio entropy). Moreover, there are many entropy changes, some very large (such as between positions XV and XVIII), that do not coincide with any of the 8 boundaries that Ferrand et al. are using as segmentation boundaries.

8.6.4. Discussion of aspects of Ferrand et al.'s theories, methods and results relevant to the current work

Ferrand et al's studies are of relevance for several reasons: They tackle segmentation from two perspectives - experimental and computational. The experimental approach used is very similar and therefore comparable to the one used in the current study in that listeners are asked to segment music in real time (chapter 3). Moreover, they suggest a set of cues that have been often discussed and provide results that enable the comparison between the location of different cues and the experimental segmentation results. They describe their work as a probabilistic (memory-based) technique, which is an inviting approach and suggests that expectation may be important in segmentation.

Probabilities, entropy and expectations

One of the central aims of Ferrand et al's study is to develop a probabilistic model expressed through the change in entropy and change in entropy level marks a segmentation boundary. This reliance on continuation (or similarity) as opposed to change is reminiscent of the Gestalt factor of similarity. The way the probability is expressed therefore, is that once a repeated pattern is established, it is expected to continue (the probability is high that it would continue). Therefore, if something with a low probability (unexpected) occurs, there is a segment boundary at that position; if something occurs that was not expected, it generates a segmentation boundary.

Ferrand et al. state that 'expectations associated with intra-opus musical information provide strong hints for segmentation points within a piece' (2002, p. 1). It seems therefore, that in Ferrand et al.'s studies, expectation is important in the sense of thwarted expectations: the expectation is for the *status quo* to remain.

This is not the central way of approaching expectation in the current study (chapters 10-12).

The musical cues

It seems from both of Ferrand et al.'s studies (2002, 2003) that the model was developed for monophonic music. This limits the number of cues that are directly observable and unambiguously described.

This approach presents a method and implementation for the systematic identification the positions of the above-mentioned cues. It allows the immediate processing of monophonic music and comparison between its results and experimental ones, allowing the comparison of the presence of these cues and the segmentations produced in, for example, listeners' experiments.

Note pitch and length (particularly relative pitch intervals and note lengths) have been suggested in many theories as the basic cues that are relevant in musical segmentation and grouping (e.g. Lerdahl and Jackendoff, 1983, Temperley, 2001, Deliège, 1998). It seems from Ferrand et al's results that, in some cases, these cues coincide with segment boundaries identified by listeners. In other cases, segment boundaries do not occur when these cues are present and in yet other cases segment boundaries do occur when at least some if not all of the above-mentioned cues are missing. It may be, therefore, that listeners are using these cues in the segmentation tasks. However, it also seems that there may be other cues that contribute to segmentation in positions where the cues investigated here do not occur (at least in *Die alte Weise* there are 13 further segmentation points that were not considered in Ferrand et al.'s study). There may also be other cues overriding those investigated here in positions where the cues occur, but there is no segmentation by listeners.

8.7 Temperley's Phrase Structure Preference Rules

8.7.1 Temperley's approach to phrasing, the Phrase Structure Preference Rules (PSPRs)
8.7.2 Application of the PSPRs to the current case-study pieces and comparison with the current MIDI results
8.7.3 Summary of Temperley's results and findings relevant to the current work

8.7.1 Temperley's approach to phrasing, the Phrase Structure Preference Rules (PSPRs)

Temperley (2001, Chapter 3, Melodic Phrase Structure, pp. 55-83)[76] presents an approach to melodic phrase structure that aims to identify phrases in monophonic folk songs. These Phrase Structure Preference Rules (henceforth PSPRs) form one module within a larger group (including metric and harmonic analysis) that is intended to model the cognition of basic musical structures. The aim of the PSPRs is to identify phrase starts. Temperley regards music from the 'common practice' tradition or other polyphonic music as too complex, its analysis resulting in ambiguity and conflict, and therefore unsuitable for computational modelling at this time (p. 65). Instead the rules are based on monophonic folk songs from the folk song collections (Essen Folk Song Collection and pieces from *Music for Sight Singing,* Ottman, 1986).

The phrase is presented as one level of grouping, which 'seems to be the level that is clearest and least ambiguous in perception (though it too is sometimes susceptible to ambiguity)' (p. 66). Temperley, in his discussion of Lerdahl and Jackendoff's work explains that the term phrase 'generally refers to basic, low-level groups, containing a few measures and a handful of notes' (p. 56). This seems to contradict Lerdahl and Jackendoff's description (see chapter 8.2) which suggested that in their grouping structure, the phrase level was relatively high.

Moreover, Temperley's description of the phrase is that it seems to be the clearest and least ambiguous perceptual entity (or unit). Though he does not include immediate musical characteristics associated with the term, the PSPRs, do define the musical criteria that he views as most important for phrase boundary identification.

The PSPRs

PSPR 1 Gap Rule. Prefer to locate phrase boundaries at (a) large interonset intervals [IOIs] and (b) large offset-to-onset intervals [OOIs] (p. 68).

---

[76] In this chapter, this text is referred to with page numbers only.

Temperley gives a formula for the relative strength of potential boundaries comparing the note length of the current note with the average length of preceding notes. In this rule, Temperley combines the ideas from Lerdahl and Jackendoff's GPRs 2a and b (slur/rest and attack point, Lerdahl and Jackendoff, 1987, p. 344) into one rule.

However, there are differences: Temperley, in calculating the 'gap score', distinguishes between the importance of IOIs and OOIs: 'A value that works well is simply to sum the note part of the IOI and two times the rest part [OOI]' (p. 68), indicating that a rest is more important than a long note of the same length. Temperley's rule looks only for long notes regardless of the note-lengths on either side. Lerdahl and Jackendoff's GPRs 2a and 2b look for a note that is long in comparison to those on either side. Referring to Tenney and Polansky's (1980) approach that 'a gap is a phrase boundary if it is larger than the gaps on either side' (p. 69), Temperley explains that this kind of rule leads to too many phrase boundaries and so instead prefers to remedy the problem by going on the next PSPR. In addition, this rule does not include the other temporal pattern in Lerdahl and Jackendoff's the change rule GPR 3d (1987, p. 344) that does not allow alternating long and short notes.

PSPR 2 Phrase Length Rule. Prefer phrases to have roughly eight notes

PSPR 2 'takes advantage of a striking statistical fact: phrases are rather consistent in terms of their number of notes' (p. 69). Temperley suggests that the range (between 6 and 10, or 4–14) may correspond to vocal breath or information processing constraints. Having studied the Ottman 1986 corpus on which Temperley developed his model, he concludes that the phrase is, on average, 8 notes long. Temperley gives a formula calculating penalties for phrases that differ from this length (p. 72).

PSPR 3 Metrical Parallelism Rule. Prefer to begin successive groups at parallel points in the metrical structure.

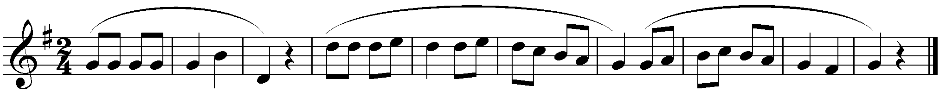
PSPR3 generalises the idea of parallelism (developed by Cambouropoulos, chapter 8.5) to look for phrase starts at parallel positions to the preceding phrase start in the metrical structure.

The main aim of this discussion of Temperley's work is to compare the results of the current MIDI listening study and his model, and see what can be learned from his model about phrase perception. Temperley explains that his model was developed on the basis of, and is only applicable for, folk music. With the initial phrase-length settings (PSPR 2) the program is only applicable to the data set taken from the Ottman book on which he trained his model. It may, therefore, be helpful to first look at the rules in the context for which they were intended before discussing them in the context of the current, case-study pieces, which are different.

Discussion of one of Temperley's examples

Temperley's first example is from the Ottman songs; Melody 103 (Hungary), (his Fig. 3.7A, p. 67, Ottman, 1986, replicated here Figure 8.7.1) and is annotated by Ottman in an 'intuitive and natural way' (p. 67).

Figure 8.7.1



Having applied **PSPR1** to this example, Temperley explains that this rule identified the correct boundaries. However, he points out an exception in bar 5 where it identifies a boundary after the first beat of the bar. This is a phrase boundary that he had described earlier as 'dubious' (p. 67). However, from his presentation and discussion of **PSPR1** it seems that the two crotchets in bar 2 and bar 9 would also be identified as potential phrase boundaries. Temperley suggests various options to remedy the situation and chooses **PSPR2.**

PSPR2 would certainly penalise the short phrases of five notes in bars 1-2 and, one note in bars 2 and 3 while the whole phrase of bars 1-3, having 7 notes which coincides with Ottman's structure, would win out. This probably holds true for bar 9, as the program assumes a phrase boundary on the last note, so would penalise a 2 and a 3 note phrase and favour the 9 note phrase also identified by Ottman. **PSPR2** would probably also penalise the 5 note phrase identified by **PSPR1** in bars $4\text{-}5^1$, however, it would probably penalise even more the 12 note phrase (bb. $4\text{-}7^1$) (being further away from the ideal 8 notes phrase than 5).

PSPR3 on the other hand would favour phrases that start on the first beat of the bar (level 3) or hyper-bar (level 4). Looking at the bar-level, it would seem that the first beat of bar 5, having been identified by **PSPR1** as a phrase start would also be identified by **PSPR3**. Looking at the hyper-bar level, if they are two-bar hyper-bars, this position would be identified again. Having had a three-bar first phrase however, it seems that the program is more likely to choose three-bar hyper bars as the more likely structure. In this case, **PSPR3** would identify the first beat of bar 7 as a phrase start. However this position is not the one chosen in Ottman's 'intuitive and natural' annotation.

It is unclear how the program would identify the 'intuitive and natural' phrase positions, and only these. The main problem seems to be distinguishing between the criteria governing the labelling of D of bar 5 which is not a phrase boundary and the G in bar 7 which is as annotated by Ottman. Harmonically, the D is a continuation of the dominant established at the beginning of the phrase (bar 4), nothing has happened yet harmonically or melodically and no harmonic resting point has been reached. This is followed in bar 6 by a 5-4-3-2-1 melodic descent which, both on its own, and because of the implied harmonic progression prepares for, and resolves in, a cadence. This is the preparation for and conclusion of the phrase on bar 7. It seems therefore, that the difficulties encountered in this 9-bar folk song may be avoided if the harmonic (or even melodic) pattern within its tonal setting, are considered.

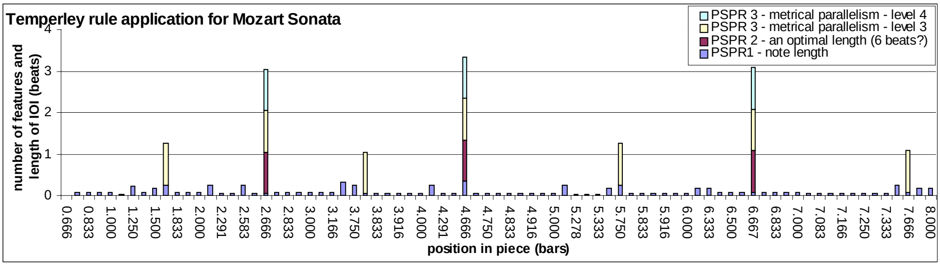
8.7.2 Application of the PSPRs to the current case-study pieces and comparison with the current MIDI results

To explore Temperley's rules in a broader musical context, they were applied to the current case-study pieces. His rules were not developed for this kind of music and were implemented in a fully developed program which could not be used here. This application is not a 'test' of the rules but rather an exploration of the applicability of the musical characteristics it uses. The program was developed for a specific corpus and needs a large corpus for the values of PSPR2 to be learned and used successfully. The current lack of a large corpus of the 'common practice' music already annotated with 'phrase marks' in MIDI or any other form means that running the program on a large number of examples is difficult. Therefore, it would not be possible to use PSPR2 though the average number of notes per phrase will be discussed in retrospect. The two remaining rules are applied manually.

There are several more ways in which these results could differ from those generated by the program. Temperley's program uses a 'dynamic programming' system of proposing several different possibilities and choosing the one with the highest score (the one that satisfies most of the rules most of the time). To some extent, although the decisions are made 'online' note by note, the final presentation is the best fit for the whole piece viewed from its end. In this exploration the rules will be applied once only. Following Temperley's explanation that the program prefers the first note of the piece to be the beginning of the first phrase and that subsequent phrases should begin in parallel positions, this is the way the rules will be applied below. Also in this application of PSPR1 the note length is not divided by previous average note length, but presented on a variable scale in relative MIDI note-length. The metrical parallelism rule PSPR3 is applied twice – at levels 3 (tactus/bar) and 4 (hyper-bar) (p.72) and is applied in the form of a binary yes/no rule.

Each graph of Temperley's rules 'application' to the case-study pieces below is compared with MIDI listener responses because, as in the case of Lerdahl and Jackendoff, this rendition is the closest to the 'input' to the program. In the following examples, the IOIs and OOIs are given at the end of the actual note; i.e. at the potential phrase start. The MIDI response graphs are in appendix 3.6

Mozart Sonata
Graph 8.7.2.1



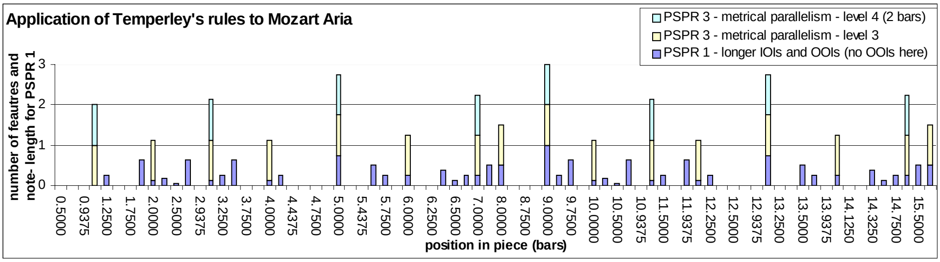
Graph 8.7.2.1 shows that the strong feature is the occurrence of the relatively long IOIs in the area of (preceding, or at the same position as), the metrical parallelisms (bars 1, 4, 5 and 7). The strongest peak (where all three features coincide) is bar 4.666. The other two peaks of the coincidence of levels 3 and 4 of the metrical parallelism rule do not coincide with relatively long IOIs. Of the four remaining level 3 metrical parallelism positions, two coincide with relatively long notes and two coincide with relatively short notes that follow long notes.

In comparison with the PS responses (graph 3.6.3.1, appendix 3.6), the level 4 hypermetrical parallelism rule is the only one that coincides only with high response for phrase starts and vice versa (high responses occur only with the level 4). This is not to say that the other rules do not also sometimes coincide with the high responses (level 3 PSPR3 does by default for half of these). However, they seem less reliable in identifying only phrase start positions. In all cases, there is a slight spread of responses around an area rather than a specific note.

The number of notes within each 'phrase' identified in response range from 15 in the first phrase to 20 in the second, 22 in the third and 18 in the last. However, if counted in crotchet beats, there is the same number of beats (6) per phrase. This is a particularly clear example of a case in which the underlying structure (for example harmonic and thematic) remains very similar but the 'surface' is varied resulting in large changes in number of notes. This information was therefore also added to graph 8.7.2.1. Indeed, Temperley mentions that the number of notes rule may not be useful for music from the 'common practice' style (pp. 82-3).

Looking again at Melody 103 discussed above, the phrases annotated by Ottman contain 7, 12 and 9 notes. If counted in crotchet beats the phrase lengths are 5, 7 and 6 beats: a much more similar 'phrase length'. Temperley excludes rests from his phrases (e.g. p. 71). If rests are included, the phrase lengths are even closer 6, 7 and 7 or 6 beats. If the aim is to find a measure of length of phrase that reflects the idea of a consistent length, the beat may be a more useful one, not only for music from the common practice but maybe also for folk music – though this would need to be tested on a larger corpus of music.

Mozart Aria
Graph 8.7.2.2

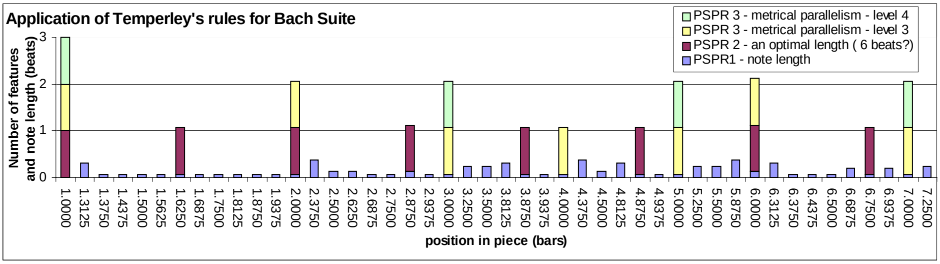


Graph 8.7.2.2 shows that in the Mozart Aria some of the levels 3 and 4 of PSPR3 hypermetrical parallelism coincide with long IOIs or OOIs in the melody while others do not. Similarly, some long IOIs and OOIs occur at positions that do not coincide with levels 3 (and 4) hypermetrical parallelism. Temperley's program would eventually reach a 'best' fit.

Listeners were responding to the whole texture of melody and accompaniment, while the rules only apply to the melody. This only affects PSPR1 as the metrical parallelisms are superimposed on the bar structure. Graphs 8.7.2.2 and 3.6.4.1, appendix 3.6 show that the main peaks of the metrical and long note rules coincide with peaks or areas of boundary identified in the responses.

In terms of number of notes, the first phrase is 12 notes and the second is 10. Both are then exactly repeated. In both cases, the phrases are 16 beats long (including rests).

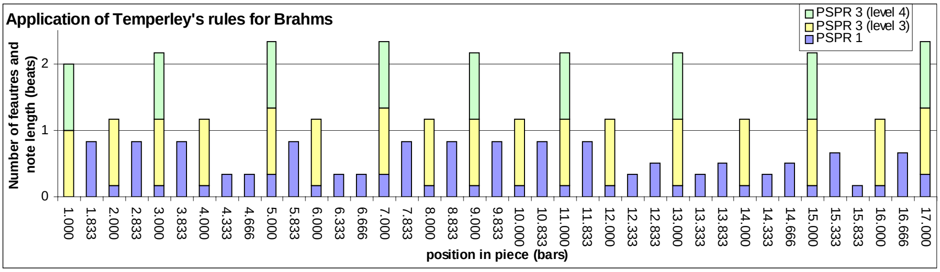
Bach Suite
Graph 8.7.2.3



Graph 8.7.2.3 shows that, unlike in the previous pieces, none of the longest IOIs or OOIs coincide with levels 3 and 4 of PSPR3 metrical or hypermetrical parallelism. However, in all cases, a long IOI follows the parallelism. Again, Temperley's algorithm would probably find a best fit. In terms of an intial application of Temperley's rules, the peaks in the rules graph coincide with areas of response in the response graph (graph 3.6.2.1, appendix 3.6). The number of notes per phrase as indicated by the responses (see discussion of responses above, e.g. Lerdahl and Jackendoff, and below) is 12, 10, 12 and 12. In terms of number

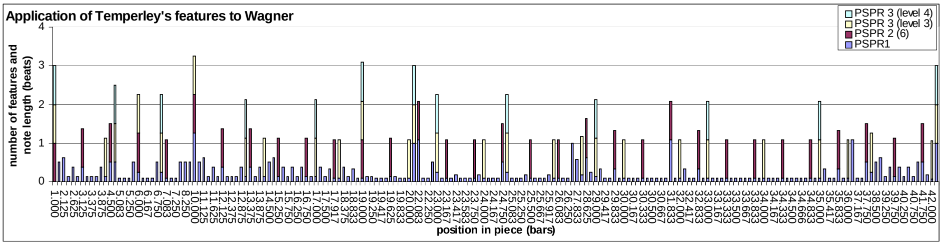
of beats, the phrase lengths are 4, 6, 8 and 8. In this case then, phrase length 'equality' is represented more by number of notes than number of beats. Similar observations are seen for the Bach Passion.

Brahms
Graph 8.7.2.4



Graph 8.7.2.4 shows that, in all cases, the longest IOIs do not coincide with the metrical parallelism rule. In comparison with the responses (graph 3.6.5.1, appendix 3.6), only the metrical parallelism rules coincide with peaks in response and response peaks only occur in coincidence with metrical parallelism rule. However, there are many metrically parallel positions, which do not have peaks in the response. The shortest phrases (in number of notes) identified by listeners is 9, 5, 4, 11, 10 and in beats 12, 6, 6, 14, 11. As will be discussed below, the second and third phrases may be considered as 1 (so: 9, 9, 11, 10 and in beats, 12, 12, 14, 11) (chapter 10).

Wagner
Graph 8.7.2.5



The comparison of graph 8.7.2.5 and the MIDI PS responses (graph 3.6.6.1, appendix 3.6) shows that again there are some positions of coincidence between long IOIs and metrically parallel positions. Moreover, in many of the cases, there is a coincidence between peaks in the Temperley rules and areas of response.

8.7.3 Summary of Temperley's results and findings relevant to the current work

The above examples indicate that there is some coincidence between metrically parallel positions and relatively long IOIs or OOIs. Moreover, there is sometimes coincidence between these features and PSs identified in response to the MIDI renditions. A comparison between the relative reliability of these and other features will be made in chapters 10-11.

Number of notes per phrase

In some cases, there seems to be regularity in phrase length in terms of number of notes. Overall, however, the variability in number of notes per phrase within and between pieces is large. The success rate of Temperley's model is currently sensitive to relatively small changes in the ideal phrase length in number of notes: Temperley changes the setting from 8 to 10 notes per phrase from the Ottman to the Essen Folk song collection. It seems, therefore, that it would be problematic to use the number of notes per phrase for a corpus of 'common practice' music. A larger annotated corpus is necessary for this to be fully investigated. The expectation for a particular length of phrase will be returned to in chapters 10-12. However, rather than a regularity in number of notes per phrase, even within pieces (especially the Mozart) the regularity seems to be in terms of beats per phrase. For example, in the Mozart Sonata each phrase is about 6 beats long but the number of notes is different in the four phrases.

Metrical parallelism

The discussion above suggested that the phrase starts identified by MIDI listeners often coincide with positions that are metrically parallel to the opening. However, there are also metrically parallel positions that do not coincide with phrase start responses and there are some phrase starts that do not coincide with metrically parallel positions as defined in the rules.

Both metrical parallelism and number of notes per phrase are template features and could reflect a theory that we learn to expect phrase starts at particular positions or because of the amount of information that has passed. They could also reflect the theory that the amount of information that we put together in a phrase is limited by memory capabilities. As can be seen from the results above the variability of phrase length seems large and so the amount of information we can keep in our memory may not be best represented in this way. On the other hand, phrase starts often did occur in similar metrical positions though not at every bar or hyper-bar in the piece. These template features will be returned to in chapters 7-9.

Long IOIs and OOIs

As with the metrical parallelism, there are some situations in which long IOIs or OOIs coincide with phrase start responses, and others in which they do not. Similarly, there are some positions in which the phrase start responses do not coincide with long IOIs or OOIs. As can be seen from the discussion in previous sections in this chapter, long notes and temporal gaps have been often proposed as being important in phrase boundary detection. This form of identifying phrase boundaries seems relatively simplified in comparison to previous approaches.

Temperley excludes rests from phrases because, he explains, phrases are often marked that way in the score. Temperley may here be referring to the Ottman and Essen Folk collections as, in much 'common practice' music, phrases are not explicitly marked at all. Moreover, most of the time Temperley seems to be modelling cognition and seems to be assuming a listener rather than score. It seems that he is saying that the listener lets the rests hang in between phrases. This contradicts Lerdahl and Jackendoff (discussed above) for whom there are no gaps between phrases.

Distinction between phrase start and end

Temperley is one of the few theorists discussed here to distinguish between phrase ends and starts. This is partly necessary because of the exclusion of rests from phrases; if there is a rest, a phrase has to end before the rest and start after it. However, his program only identifies phrase starts and the features directly connected with them. It may be important to also identify phrase ends and the features related to them.

Specific phrase starts and ends or areas and expectations of them

Temperley explains that our intuitions about phrase structure are often 'indecisive' (p. 83). As can be seen from the responses in the listening experiments there is indeed a variety of responses when looking at specific locations for phrase boundaries. However, as suggested in chapters 3 and 4, looking more at areas rather than specific positions, and the features that relate to phrase boundary expectations in certain areas, may lead to the identification of more commonalities between listeners, possibly reflecting clearer 'intuitions' about these aspects of phrase perception.

Preference rules and variable weighting

These and other preference rules (including Lerdahl and Jackendoff), allow the possibility of proposing different phrase starts and choosing the best structure among the options. In addition, having a varied weighting of some features dependent on, for example, their length, allows a relatively fine-grained approach to different features.

Rule base and memory-base

Even though it seemed that Temperley's program is based on a rule base, one third of the rules (PSPR2) seems to function more as a memory base. The 'setting' for the rule for average number of notes per phrase needs to be learnt from each corpus. Even within similar music (such as songs from the Ottman and Essen Folk Song collection) there seems to be a dramatic difference in success depending on whether or not the program has been trained on one or the other corpus.

Such a combination of a rule base and a memory base may be the most useful in identifying the rules and features that affect our perception of phrases. Unfortunately, currently there is no equivalent corpus for music from the 'common practice' era.

8.8 Palmer and Krumhansl's experimental approach to phrase completeness judgements. With Lerdahl & Jackendoff metre and time-span reduction and Krumhansl and Kessler's ratings of harmonic fit models

8.8.1. The contribution of temporal and pitch information to phrase completeness decisions in melodies
8.8.2. The applicability of the theoretical models for phrase perception.
8.8.3. Applying Krumhansl and Kessler's model to the current case-study pieces as described in Palmer and Krumhansl
8.8.4 Summary of Palmer and Krumhansl's results and findings relevant to the current work

8.8.1. The contribution of temporal and pitch information to phrase completeness decisions in melodies

Palmer and Krumhansl (1987a) study the contribution of temporal information (relative note lengths, analysed as metrical structure) and pitch structure (relative pitch height, analysed within a harmonic framework) to phrase completeness decisions in melodies (see also Palmer and Krumhansl 1987b). Two aspects are of particular interest for the current study: the relationship between 'temporal' and 'pitch' information as understood from Palmer and Krumhansl's experimental results, and the applicability of their theoretical models in the exploration of phrase perception.

Palmer and Krumhansl conclude that an additive model of temporal and pitch patterns is sufficient for the prediction of phrase completeness ratings. Their results suggest independent pitch and temporal factors in melodic phrase judgement. They compare their experimental results with predictions based on Lerdahl and Jackendoff's metre and time-span reduction models (1987) and Krumhansl and Kessler's ratings of harmonic fit model (1982).[77] Their test piece is Fugue XX in a minor, bars 1-4 by J.S. Bach, from the *Wohl-Temperiertes Klavier I* (figure 8.8.1.).

Figure 8.8.1

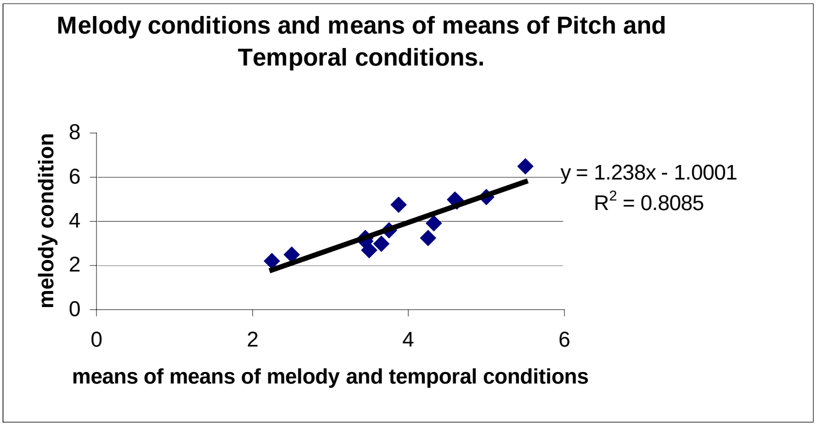
For this piece, comparison of the results with the two sets of theoretical models indicate that perceptually strong tonal events are not paired with metrically strong

[77] Krumhansl and Kessler (1982) model the stability profile of tones in a major key by measuring the perceived goodness of fit of tone added to a fragment.

beats (Palmer and Krumhansl, 1987a, p. 121) The effects of pitch and temporal events were additively combined with fairly equal weights to predict phrase judgements (1987a, p. 125).

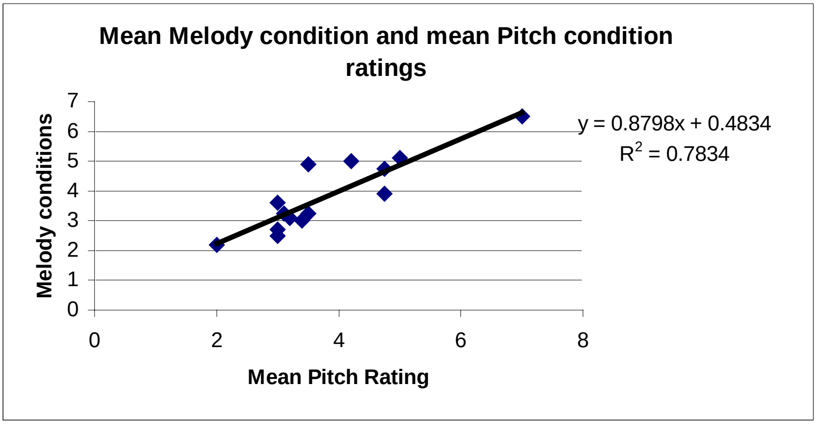
The results of the correlation between the pitch and temporal conditions on one hand, and the melody condition on the other, are plotted below (graph 8.8.1). The means were read from the graphs (presented in units of 1) given in Palmer and Krumhansl (1987, p. 120). There might be an error of 0.1 in reading the results. Moreover, small differences in response can result in very different correlation coefficients, showing that the model is quite sensitive. This is seen by comparing the plot and correlation coefficient that result from reading the data from the paper with the correlation coefficient of the same data-set in Palmer and Krumhansl's Figure 4 (1987a, p. 120). The figure for R here is slightly lower than the one given by Palmer and Krumhansl (1987a, p. 120), R=0.94 (i.e. $R^2 = 0.8836$) but as discussed, it seems that the model is very sensitive. The following discussion, however, is based on the big differences.

Graph 8.8.1



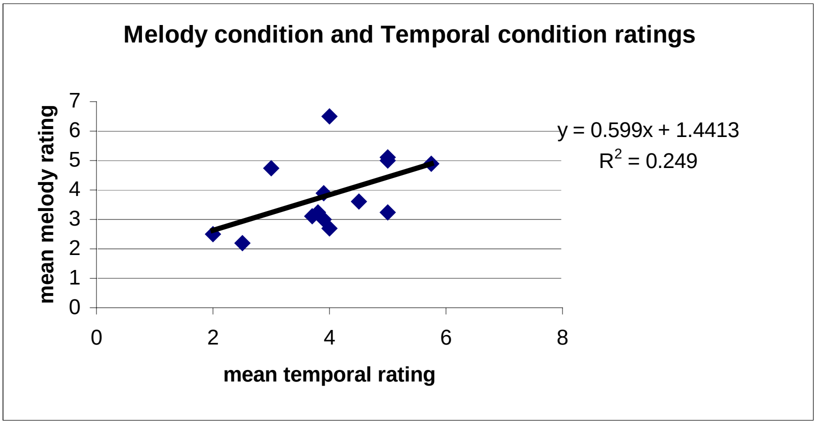
The results of the pitch and melody conditions (1987a, p. 120), re-plotted below (graph 8.8.2) as a scattergraph, show that the mean pitch ratings are highly correlated with the mean melody ratings.

Graph 8.8.2



On the other hand, the results of the temporal and melody conditions re-plotted below (graph 8.8.3) indicate a low correlation, both these correlations are lower than the correlation between the averages of temporal and pitch condition and the melody condition.

Graph 8.8.3



It is notable that there are two positions in which the correlations of pitch and melody condition reduce the overall correlation. At note 22 the rating response of temporal and melody conditions is the same whilst at note 29 the high rating of the temporal rating compensates for the low rating of the melody condition (note numbers are in figure 8.8.1).

The substantially better correlation of pitch and melody ratings in comparison with temporal and melody rating indicates that the contribution of the pitch rating is higher than the temporal rating to the overall melody rating. Furthermore, Palmer and Krumhansl explain that listeners heard the melody condition first tended to show a greater correspondence between judgements in the temporal and melody conditions than those who heard the temporal or pitch conditions first (1987a, p.121). This implies that the correlation between the results of the temporal condition and the melody condition would have been even lower if the group that had heard the melody first would have been excluded. However, this influence was not recognised in the pitch rating. Palmer and Krumhansl explain this by a possibly stronger memory effect of the pitch structure. However there may be additional explanations:

1) The correlation between the pitch condition and the melody condition is already high, so the potential for improvement is small,
2) The rhythm of the second half of the excerpt in the melody condition is almost equitemporal and is therefore similar to the equitemporal pitch condition making the difference between the two conditions rather small, and smaller than between the temporal and melody conditions.

It is not clear whether there is an influence on listeners hearing the pitch before the melody conditions. The note lengths in the first half of the excerpt are more varied and, therefore, may stimulate a perception of a metrical structure in the melody condition which will not be possible in the equitemporal condition. However, Palmer and Krumhansl state that the shortened segments may not provide enough information to establish the metrical structure of the complete excerpt (1987a, p.121).

The results of Palmer and Krumhansl's experiment using Fugue XX show that combining pitch and temporal information additively provides the highest correlation with melody condition responses. However, as the discussion above shows, pitch is more important than tempo. Therefore, an equal weight, additive model does not seem to be a good presentation of the contribution of these two factors. A closer look at the two anomalous positions in the correlation between the pitch and melody conditions reveals that the addition of temporal information to pitch improves the correlation and supports the idea of an additive model.

For note 22, the high response in the pitch condition may be due to its harmonic position as a tonic. This importance is significantly reduced by its metrical position. Even though figures for the mean rating at that pitch rating is one unit higher than the melody rating at this position, the contour of ratings reaches the highest value since the beginning of the section.

For note 29, a relatively low response to the mediant note in the pitch condition is compensated for by a high response in the temporal condition. This position, when presented with the true rhythmic pattern is parallel to the end of the previous first unmodified section of the excerpt.

This brief musical analysis of these two positions illustrates the need for detailed analyses in order to explore the reasons for such anomalies and improve the models. More specifically, this analysis showed the necessity to include the temporal factor in the prediction of phrase completeness. At the same time, however, the broader comparison showing the higher level of overall success of the prediction based on the pitch condition, indicate that pitch should be considered as a more important factor. The results leave open the possibility of assigning relative weights to these factors.

8.8.2. The applicability of the theoretical models for phrase perception.

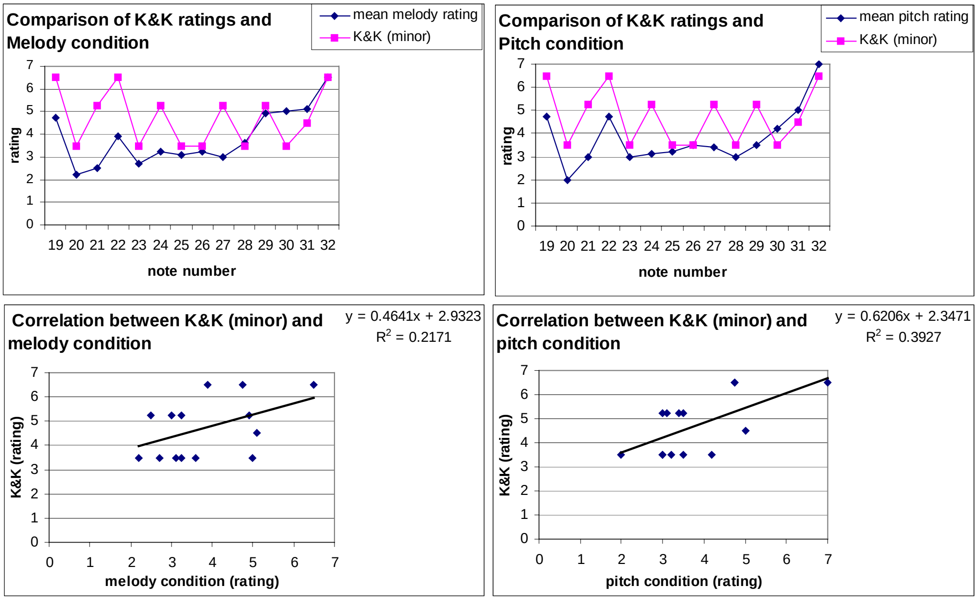
8.8.2.1 Lerdahl and Jackendoff

Of the different parts of Lerdahl and Jackendoff (1987), Palmer and Krumhansl choose to concentrate on the Metrical Preference Rules for the temporal condition and the Time-Span Reduction rules for the melody condition. However, they conclude that predictions from the Metrical Preference Rules did not correlate well with ratings for the temporal condition while the time-span reduction predictions did supply accurate predictions for melodic phrase judgements. These time-span reductions are done, according to Palmer and Krumhansl (1987a, p. 121), on the basis of Lerdahl and Jackendoff's metrical and grouping rules. Having established that the metrical preference rules did not correlate well with their results, and having claimed that the grouping rules are inapplicable to this question (1987a, p. 118), it remains unclear what additional information can be derived about the musical features contributing to the phrase completeness responses. Palmer and Krumhansl show that when metrical information is removed, the correlation is reduced and is lower than that obtained from comparing Krumhansl and Kessler with the melody condition (R =-0.58, and R = 0.61 respectively, 1987a, p. 121). Moreover, when removing metrical information from Lerdahl and Jackendoff's time-span reduction rules, it seems that the only features left are cadential and relative note-height (preferring higher notes). It is unclear exactly how these rules were applied when the metrical information was removed. Highest pitches, as feature on their own, in their test example, do not seem to coincide with high ratings in the pitch or melody conditions. Cadential features will be discussed elsewhere (chapters 10-12) and general pitch features are discussed in section 8.8.2.2. Therefore, the time-span reduction rules as applied in Palmer and Krumhansl's study, will not be applied to the test pieces of this study.

8.8.2.2 Krumhansl and Kessler

Palmer and Krumhansl (1987a) also apply Krumhansl and Kessler's (1982) ratings of pitch to their results. Krumhansl and Kessler's results are presented in Krumhansl and Kessler (1982, pp. 343-344, hereafter, K&K). The graphs are not presented in Palmer and Krumhansl (1987a). However, graphs based on those in K&K and Palmer and Krumhansl (1987a, p. 120), are provided below. Here we obtain slightly different correlation coefficients from Palmer and Krumhansl (p. 121) however, as mentioned above, the data is read from the graphs in the articles and the models are very sensitive to slight changes. Nevertheless, these graphs reveal several interesting characteristics.

Graphs 8.8.2.2



The graphs showing the pitch condition against K&K show a correlation between the two. Palmer and Krumhansl conclude that the degree completeness of a phrase on a particular note is related to the degree of pitch 'fittedness' to the key. However, despite the correlation figure the ratings of the pitch condition and K&K are not equivalent:

While there is a distinction between highly-rated (pitch A) and low-rated areas (the rest) in the pitch condition, the K&K contour seems to have three levels: high (A pitch), medium (C pitch) and low (the rest). Moreover, the graph above shows that responses to the tonic pitch differ under different conditions: notes 19 and 22 are rated the same, whilst the last note (note 32) is rated much higher. It seems, therefore, that given the same pitch, the perception of phrase completeness varies with its position.

Moreover, the graphs representing the melody condition against K&K show that the correlation is less good than in the pitch condition. This suggests that when pitch information is combined with other information, a scale of pitch 'fittedness' based on individual pitches and their relationship to fragments (scales, chords, cadences), is even less reliable for identifying phrase completeness.

The main difference between the melody contour and the K&K is, as with the pitch contour, that while there are continuous peaks and troughs throughout the K&K contour, in the melody responses these are fewer and more distant.

However, two characteristics may shed light on the relationship between pitch 'fittedness' information and the phrase completeness responses:

Firstly, the response to the tonic is different in different positions. The difference here is larger than already observed for the pitch responses above. Now both the responses to notes 22 and 19 also differ: note 22 has the lowest response, note 19 a higher response and the last note (note 32) the highest response in the piece. This not only indicates that the same note (with the same pitch 'fittedness' rating) elicits a different response in different positions, but also that when combined with 'temporal' information, the same note can also elicit different responses. This may be because the notes are of different length; notes 19 and 32 are quavers, while note 22 is a semiquaver. If this were the case, we may expect the responses to notes 19 and 32 to be similar (though both are preceded by different note lengths which may cause differences in response). Another explanation is that all three occur on different metrical positions, note 19 is an upbeat (weak), 22 is on the third semiquaver of the first beat (weaker) and the last is a down beat (strongest).

Secondly, the responses to the melody of notes 29 – 32 are different in these positions from the same notes in other positions, while in the K&K contour they are the same. The difference is also seen in the pitch response contour (above) but here it is more pronounced. For example, the response to notes 29 and 30 is different from the response to notes 20 and 21 (there is no other E, note 31, for comparison) in which the note order is different. Notes 29-31 are also the only quavers with these note-pitches and, as discussed above, note 21 is in the same metrical position as the parallel in the previous section. This suggests that the pitch context (or order) affects the phrase-completeness responses. In addition, as can be seen from the bigger difference in the melody condition as opposed to the pitch condition, the temporal context may affect the phrase-completeness responses even more. It remains unclear from this example, however, what the precise reasons are for the differences in responses.

In more general terms, there seems to be coincidence between the K&K contour and the responses to the melody condition in terms of the highest peaks. All the tonic–note peaks in the K&K contour coincide with local peaks in the melody response contour. It seems therefore, that only tonic-note pitches are candidates for complete phrase judgements. This implies that phrase ends are perceived only on the tonic note, but that not all tonic notes are phrase ends. The more fine-grained implications of this conclusion, such as the identification of the 'correct' tonic pitches that coincide with phrase ends, and an exploration of more general applicability is difficult to carry out on the basis of this data because the study of Palmer and Krumhansl deals with only a bar and a beat of one musical example. Therefore, the K&K empirical grading of tone-fittedness is applied to the current case-study pieces followed by a comparison with my MIDI listening response data.

8.8.3. Applying K&K to my test pieces as described by Palmer and Krumhansl

If it were indeed found that there was a systematic relationship between the K&K ranking and phrase boundaries, then the above observations would be most helpful in phrase end (PE) identification, particularly as Palmer and Krumhansl discuss phrase completeness. The K&K rankings were therefore applied to the current case-study pieces.

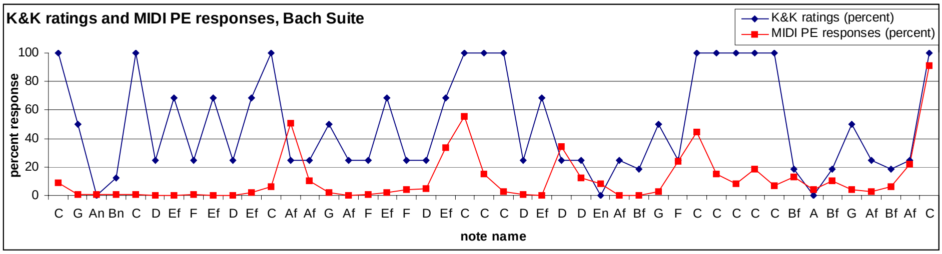
Palmer and Krumhansl use artificially generated examples with note lengths being assigned according to the score. Therefore, the MIDI listening results are used in the discussion below. Unlike the Palmer and Krumhansl study, however, listeners did not give rankings for each position. Instead, they indicated phrase starts and ends and responses are recorded as percentage of total listeners (chapter 3). The highest peaks are 'stronger' PEs identified by the majority while lower peaks are 'weaker' PEs identified by only some listeners. One of the main conclusions from the above discussion was that the tonic notes should coincide with (and predict) candidates for PEs.

In K&K's experiments, they presented subjects with a 7–level scale to identify pitch-fittedness. However, the mean responses only cover the range 2.5 to 6.5. To make the scale used by K&K and the percentage response in the MIDI experiment comparable, the K&K rankings were scaled up by subtracting the lowest (background) rank (2.5) and expanded the range to 100 % (multiplying the value by 100/4, 4 being the actual range of K&K rankings). In order to use the most similar pieces, the three case-study pieces with the clearest single melody lines (the Bach Suite, Mozart Sonata and Wagner) are used in the following comparison.

8.8.3.1 Bach Suite

The Bach Suite is most similar to Fugue XX; it is mostly monophonic and by the same composer. The contours of the K&K profile and the MIDI responses (graph 8.8.3.1) are similar to those found for the Bach fugue excerpt. The K&K contour has several peaks at different levels, while the MIDI responses have fewer peaks that are further apart. Moreover, as graph 8.8.3.1 of listener MIDI PE responses and the K&K profile shows, three of the five peaks of MIDI listener response coincide with the tonic note. Of the two remaining peaks one immediately follows a tonic note and the other follows a mediant note (i.e. tonic of the relative major, ranked one level lower than the tonic).

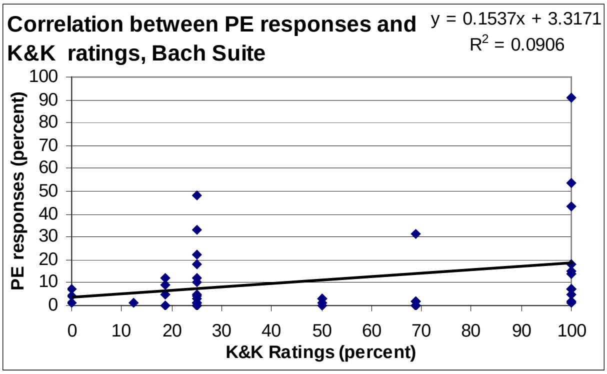
Graph 8.8.3.1



Two of the peaks are with the first of several consecutive tonic notes (bars 3 and 5, the first note of the piece is also the tonic). There are also two tonic notes that do not coincide with peaks, one (discussed above) has a peak in MIDI response one beat later, and the other does not coincide with a peak at all (bar 1.5). It seems, therefore, that though some tonic notes coincide with peaks in the MIDI listeners' responses, tonic notes do note always coincide exactly with PEs and PEs do not always coincide with tonics. Therefore, tonics alone do not give a clear prediction of the PE response. Moreover, as observed for Fugue XX, the correlation between note ranking and percentage PE is very low. The two PE positions that do not coincide exactly with tonics coincide with scale degrees that are ranked at the fourth level down the ranking of scale pitches (sub-mediant, degree 6, and supertonic, degree 2). Therefore, the position of a note within the K&K ranking does not seem to predict the relative likelihood of a PE.

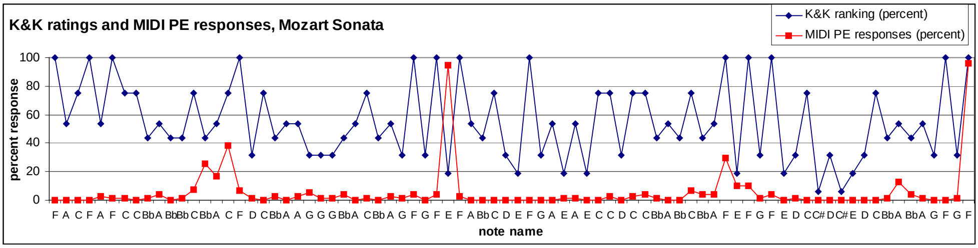
These results indicate that the pitches that coincide with phrase boundaries are mainly tonics (and one sub-mediant and supertonic). The lack of correlation (graph 8.8.3.2), the observation that the only pitches to coincide with PEs that are not the tonic are far down the ranking scale, and the occurrence of tonics not at PEs, suggest that the K&K pitch profile cannot be used for the prediction of PEs.

Graph 8.8.3.2

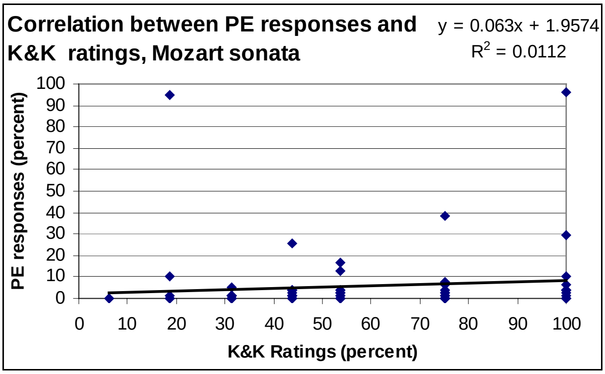


Similarly, for the Mozart sonata (melody line only) and the Wagner, the correlation between the K&K ranking and the MIDI listener results is low. In both cases the highest response is still to the tonic note (graphs 8.8.3.3-8.8.3.6).

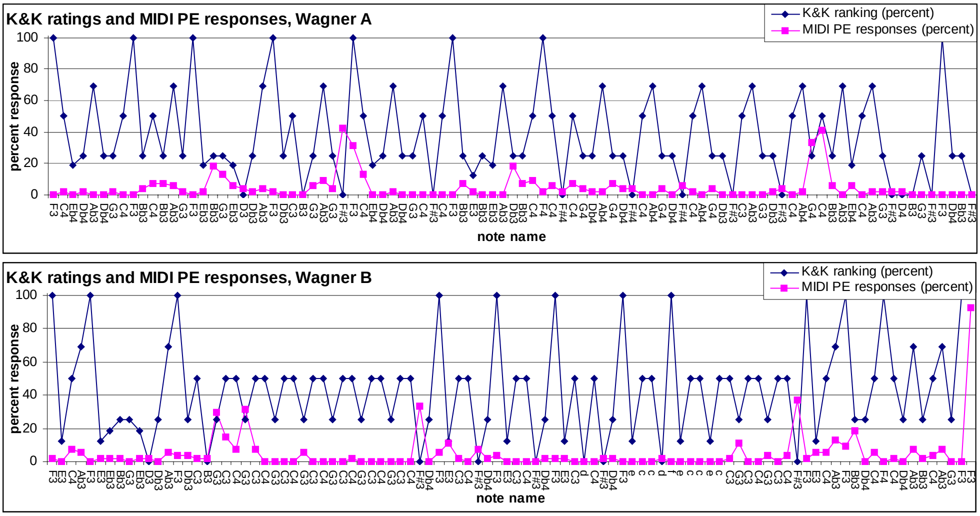
Graph 8.8.3.3



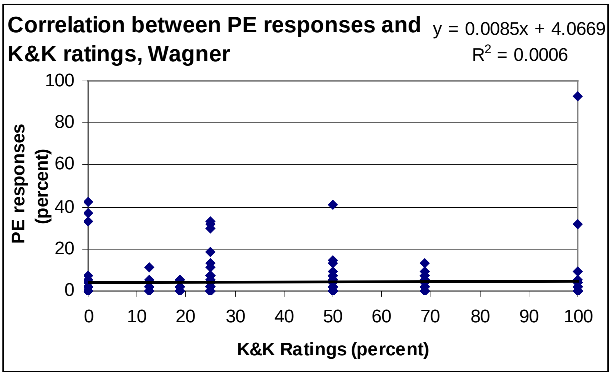
Graph 8.8.3.4



Graph 8.8.3.5



Graph 8.8.3.6



8.8.4 Summary of Palmer and Krumhansl's results and findings relevant to the current work

Palmer and Krumhansl's study explored both the cues that contribute to the perception of phrase-completeness and models that describe them. In terms of the cues, their results suggest that the relationship between pitch and temporal information is uneven in that pitch information seems to provide more information than temporal information most of the time. However, in some positions, temporal information is also necessary. It seems, therefore, that both cues contribute to phrase perception possibly in an additive but unequal way.

In terms of the models tested, it that of Krumhansl and Kessler (1982) seemed most successful in predicting the listener responses. However, on closer inspection, it the correlations between the K&K rankings and both Palmer and Krumhansl's responses and the MIDI listening responses seem low. However, it may be that the functions of the notes within the scale may be used as a guide to phrase endings and, moreover, that taking into account the context of the appearance of individual notes may help.

8.9 Summary

8.9.1 Introduction

The above discussions of theories, experimental studies and models, and the comparisons with results of the current listening studies, allowed the exploration of several aspects of current considerations of phrasing and phrase perception. In this summary, specific aspects of these theories that are considered key for this study are discussed in more general terms. This summary ends with the outline of the directions that will be followed in the subsequent chapters.

8.9.2 Definition of 'phrase'

The term 'phrase' is mentioned and described to different extents in each of the studies discussed, however only a couple refer specifically to this term and investigate its properties (Palmer and Krumhansl, 1987a; Temperley, 2001) while the others mention it more loosely. Repeatedly, the phrase is regarded as a type of group or chunk. It is also often referred to as an entity which is different from other groups or chunks. The clarity of the distinction in definition-terminology, in properties between the phrases and other groups or chunks, and the ways to tell them apart vary among the theories. Moreover, the distinction is usually implied rather than explicit. These musical groups or chunks have been adopted from the psychological literature, primarily based on the gestalt principles, and in the new musical context they are used as rather general terms. The concentration on musical instantiations of these psychological cues, primarily codified in the Gestalt principles used in 'structuring' (grouping, chunking, phrasing) the music by the listener, results in a definition of the phrase, both musically and psychologically primarily by its boundaries (section 8.9.3).

The general meanings of the term phrase are not related to a specific musical context in any of the studies discussed above. There are no general musical definitions or explanations until the rules or specific musical examples are described, usually already within the technical steps of phrase-identification. This may be a result of the consideration of these entities as general psychological ones that happen to be instantiated in a musical context. In this case, a specific definition pertaining to the musical context seems not to have been regarded as necessary. Possibly, the absence of such a definition leads to the absence of the respective musical notions and terms from the vocabulary of terminology, approaches and theories.

The combination of results and conclusions of previous studies with those of the new combined approach followed here will lead to a synthesis of a number of different perspectives and ways of understanding and defining the musical phrase.

8.9.3 Phrase boundaries

Phrase boundaries and methods of their identification

Many of the studies discussed above concentrate on boundaries, whether of groups, chunks or phrases, and specifically on their identification using musical cues. The decisions as to the location of the position of phrase boundaries seem to be mainly in the form of binary yes/no decisions, though different cues or combination of cues may result in a scale of certainty and strength of the perceived boundary.

Different theories use different time-windows for the identification of phrase boundaries. For example, Lerdahl and Jackendoff, for the initial identification of candidate positions for phrase boundaries use a short time-span of four notes (for example GPRs 2 and 3, Lerdahl and Jackendoff, 1987). In addition, or instead, longer time-spans can be used. Sometimes this involves a retrospective assignment of the phrase boundary, which necessitates backward shifting of the time window (Lerdahl and Jackendoff, 1987). Sometimes, a limited time-span is not considered but instead a predetermined 'template' for the search for the phrase boundary can be superimposed (for example, Lerdahl and Jackendoff, 1987 GPRs 4ff; Temperley, 2001).

In summary, the above discussion of the previous approaches shows that phrasing is represented usually as a boundary identification task; the internal structure of the phrases is often disregarded and interpreted as irrelevant for phrase identification. One exception is the approach of Bod (2002), which aims to identify in new pieces whole phrases that occurred previously in an annotated corpus.

This study explores the importance of "boundaries" in phrase perception and, moreover, discovers the variety of phrase-parts and their important rôles in phrasing (both in phrase construction and perception).

Phrase boundaries vs. phrase starts and phrase ends

Not only do most of the studies described above concentrate on phrase "boundaries", separating each phrase from its neighbour, they identify only one exact boundary position. Most studies consider phrase-start to phrase-start, end to end, or boundary to boundary units. Temperley (2001) does distinguish between phrase starts and phrase ends, but models only phrase starts. To some extent, this is understandable. If phrases, groups, or chunks are units, it should be possible to identify boundaries between them. However, there seem to be differences between the functions of phrase ends and starts. The temporal difference between the phrase end and start is demonstrated in the examples given by the above authors themselves, particularly those they treat as difficult or special; the cases of overlapping or elided phrases. In the first case, the beginning of one phrase occurs before the end of the previous one. In elided phrases, the phrase start occurs at the same position as the previous phrase end. This case is treated as special because the boundary cannot be between notes, indicating that usually there is a boundary between two notes. In this case, the position of the boundary between notes removes either the end or the start. These cases are difficult to describe, let alone 'model', if no distinction is made between phrase ends and starts. That is not to say that there are no cases in which the phrase ends and starts are indeed heard together as one position.

For Temperley (2001) rests are spaces between phrases and he explains that this is due to the marking of scores. This approach, however, seems to add ambiguity. An opposite view is that every note and rest should be included in a phrase (for example, Lerdahl and Jackendoff, 1987).

The use of the idea of one 'boundary' position and the usual positioning of that boundary between two notes seem to be, at least in part, driven by modelling needs and disregards musical characteristics that contribute to phrase perception. So far, the distinction between phrase ends and starts here has been the temporal one. However, there is also the functional difference between them; elements that indicate a phrase end are different from those that indicate a phrase start. This study explores the differences between phrase ends and starts which together lead to the perception of phrase boundaries.

Phrase boundary positions vs. phrase boundary areas

This study explores the characteristics of responses at phrase "boundaries" and the results indicate that phrases are perceived during listening as having boundary areas rather than having specific boundary positions (chapters 3, 4, 6, 10-13). Here the term 'area' refers to two characteristics: 1) Listeners may identify different boundary areas and respond to different specific features within them. 2) Listeners may expect a phrase to end within an area, or may assign a phrase start or end in an area in retrospect rather than relate the phrase end or start automatically to one note. In both cases, listeners may not be able to identify the exact position, but they recognise the boundary area.

8.9.4 The listeners' perspective

Many of the authors discussed above investigate and describe groups, chunks or phrases from the point of view of listeners; all the experimental studies discussed above were carried out using auditory stimuli, the only non-auditory input being Deliège's time line (1998). In the non-experimental studies, when they are explicit about 'who' they aim to model, or describe, it is the listener. Some, such as Lerdahl and Jackendoff (1987) even go on to specify primarily experienced listeners.

A broader approach is taken in this study in which the listeners' phrase responses are explored alongside the investigation of phrases identified 'from the score', phrase features as defined by musical theorists, and those represented in musical performance. This not only provides a broader basis for the exploration of the phrase but also influences the understanding of the phrase from the listener's perspective.

8.9.5 Musical education and experience, and familiarity with the piece

Some of the models are developed particularly for the 'experienced listener' (Lerdahl and Jackendoff, 1987). However, most do not make explicit claims as to which level of ability or experience they are modelling. Deliège in her experiment however, does distinguish between different levels of musical experience (Deliège, 1998). She concludes that subjects from different musical backgrounds segment in a similar way and are similarly consistent. In other words, the ability of segmentation is common to all subjects independent of musical training. She also concludes that the degree of familiarity with a piece has a marginal effect on segmentation.

This study further investigates the effects of musical education and experience, musical experience with the genre explored in this study and familiarity with individual pieces, using a larger group of pieces.

8.9.6 Type of music

The music considered in the studies discussed above, is usually monophonic (or homophonic). Several of the studies concern folk-songs such as those from the Essen Folk song collection (Bod, 2002; Ferrand et al., 2002; Temperley, 2001), while others use either monophonic music (Deliège, 1998; Palmer and Krumhansl, 1987a) or music that is made monophonic i.e. taking only the melody line (Cambouropoulos, 2001; Cambouropoulos, 2003; Ferrand et al., 2002; Lerdahl and Jackendoff, 1987; Palmer and Krumhansl, 1987b). The arguments for using folk music are clear; this is music that has been developed and sung by people with no need of special musical training and not learnt from notation. Moreover, in this (and any other vocal music with a text) the text can provide other, non-musical, yet integral cues and thus possibly provide more information. However, the texts are not used in the discussion or analysis of this music in any of the studies (Cambouropoulos, 2001; 2003, Bod, 2002, see also, chapter 2.2).

Moreover, by using only monophonic music, much information, such as harmonic or textural information, is lost or omitted. Western tonal monophonic music is often viewed as containing an implied harmony that is 'heard' by performers and listeners. However, as it is not explicitly represented in the notes, and as so much else is taken as being explicitly represented (note length, metre etc.),[78] the result is that harmony is usually (though not always, see Lerdahl and Jackendoff, 1987) excluded from the discussion and explanation.

In addition, re-casting polyphonic music into a monophonic form avoids the need to identify phrases in polyphonic music. This is problematic, partly because it is unclear how the different parts in a polyphonic texture contribute to phrase perception; for example, it is unclear whether individual parts in the polyphonic texture are treated separately and thus phrased separately, whether there is a single part that always dominates our phrase perception, or whether there is a combination of the two depending on the musical context and interests of the listener. All of the theorists mention this problem, and Temperley attempts a descriptive approach to the question but does not include polyphonic music in his model.

The music explored in this study is both monophonic (originally written as such) and polyphonic. Nevertheless, even in the context of monophonic music, possible implied harmonies are suggested. When polyphonic music is discussed, the question of the contribution of different voices or of the combination of them to the identification of phrases is explored (chapter 10).

8.9.7 Musical Cues

Several musical cues have been discussed in the above studies:

1. Relatively long IOIs or OOIs
2. Changes in note-length
3. Relatively large pitch intervals
4. Pitch fittedness within tonal hierarchy
5. Metrical position in bar – parallel to opening
6. Number of notes or bars per phrase
7. Changes (register, texture, articulation, dynamic etc.)
8. Repetition

---

[78] With the understanding that a performance cannot be fully represented in the score

Discussion of the above theories and the comparison of these with the results of the current MIDI listening study emphasise that some cues - long IOIs or OOIs, changes in note-length, relatively large pitch intervals and the pitches that 'fit' best in the tonal hierarchy (1 - 4 above) – do not occur only at phrase boundaries. It seems difficult to identify a relative scale for each of these cues and, moreover, the examples above suggest that a relative scale for the individual features would be of limited use as even within a single piece two occurrences of the same cue of the same 'strength' are associated with different positions within the phrase. For example, in the folk song discussed in Temperley's study (chapter 8.7), two equally long IOIs that would be equal on a relative scale of long IOIs, are shown to have different functions in the phrase: one coincides with a phrase end (annotated by Ottman) and the other does not.

From the above discussion, it seems that cues such as metrically parallel phrase starts and equal number of notes or bars per phrase (cues 5 and 6) may be informative in some pieces for identifying specific positions. Changes and repetitions (7 and 8), though more difficult to define than the other cues, seem to be helpful for phrase identification; overall they occur more rarely but when they do occur, they coincide with MIDI listener responses or, in the examples provided by the theorists above, with the annotations of phrase boundaries. Lerdahl and Jackendoff mention tonal stability as part of their grouping rules. The way this is discussed seems to be in the same category as the other cues - as instantaneous cues (chapters 11-13).

One of the main aims of this study is to explore the different cues used in phrase identification and the relationships between them. Though the cues listed above are often mentioned in these studies and seem to feature in some way in phrase identification, they do not explain all of the annotations discussed in the literature, or the listener responses recorded in this study. Furthermore, the theories in which these cues have been described concentrate primarily on identifying boundaries and not on the process that leads to the boundary identification, so there may be other possible musical indicators that should be considered. Moreover, this study explores the possibility of viewing specific cues as representatives of larger groups, each group having a more general function (chapters 10-13).

8.9.8 Online vs. offline models

Many of the models discussed above aim to provide the best overall parse for the phrase structure of a given piece. Therefore, although the process is to go through the piece from beginning to end, note by note, the final decision about the phrase boundaries seems to be made in retrospect ('offline'). Similarly, and perhaps surprisingly, in some of the experimental approaches the decision-making is, in large part, retrospective. For example, in Deliège's phrase identification tasks, listeners had one or more familiarisation listenings and visual memory aides. Again, even though, to some extent, there is a moment-to-moment response, it seems that the responses were, to some extent, more representative of 'offline' retrospective analysis.

This study compares moment-to-moment (online) and retrospective or holistic (offline) phrase identification both in terms of the musical cues relied on, and the phrases identified. For example, it seems that while listening to the piece, especially for the first times, the emphasis seems to be more about expectations of, or surprising phrase boundaries (online). However, in retrospect (or when the piece is known, 'offline'), it seems that the memory is of the larger units, and the boundaries (phrase starts and ends) are indeed more important. These different views of the phrase identification process have repercussions for the definition(s) of 'ambiguity' (chapters 10-13).

8.9.9 Ambiguity and confusion

The term ambiguous is often used in studies about phrasing but there seem to be several different connotations of the term: 1. Computational: For some theorists while the listener's intuitions are clear as to where phrase boundaries should fall, the ambiguity is a computational problem to be solved (Bod, 2002). 2. Vague intuitions: For other theorists, the difficulty is for the listeners – our intuitions are vague as to where phrase boundaries should fall (such as Temperley, 2001).

It seems, however, that more subtle distinctions may be made:
1. While listening to a piece, especially for the first time, there may be confusion; unexpected phrase boundaries, expectation for phrase boundaries that are not fulfilled, or areas that sound like phrase boundaries until a clearer distinction follows some time later ("online confusion"). These ambiguities may be reduced when the whole piece is known, or if the score is seen while listening, allowing immediate comparison with what has already occurred and what will follow.

2. In some cases, however, even if the piece is known, distinct, different phrase boundary options may be identified by the same or by different listeners.

3. In still other cases, whether the piece is known or not, there are cases where there is confusion i.e. where no clear phrase boundary areas are identified.

In this study, an attempt is made to distinguish between the different ambiguity types and to relate them to the presence or absence of different musical features (chapters 10-13).

8.9.10 Rules

The rule based approaches, both computational (Cambouropoulos, 2001; 2003; Ferrand et al., 2002; 2003; Temperley, 2001) and unimplemented (Lerdahl and Jackendoff, 1987), and other computational methods (Bod, 2002; Ferrand et al., 2002; 2003; and parts of Temperley, 2001) all take slightly different approaches to implementing their theories of grouping, segmentation or phrasing. All assume that one favourite phrase structure is reached at the end of the piece though many acknowledge that during the piece, several different options may be possible.

In this study, possible musical features and the phrase-parts involved in phrase identification are investigated leading to the identification of a set of 'rules'. The general form of the rules is; if musical features are present, they may indicate certain phrase parts at positions or over areas. These different musical features and phrase parts and their combinations may then indicate phrase types. There is no assumption that one phrase interpretation should "win out".

8.9.11 Further investigation of aspects of phrasing in the rest of this study

In what follows, the above matters are explored in more detail with the aim of:
1) Identifying components of a phrase definition.
2) Identifying the features that contribute to phrase identification.
3) Exploring further the musical features that occur within phrases, rather than only those at the boundaries.
4) Exploring the approach of seeing phrases being bounded by phrase starts and phrase ends, and those boundaries being areas rather than single notes or positions between notes.
5) Exploring further the different 'modes' (online and offline) of phrase perception,
6) Exploring the influence of musical education in general and familiarity with a piece on phrase perceptions.
7) Exploring extra-musical considerations (chapters 10-13).

Some of the results of these explorations are summarised in the rule base of chapter 14 and tested in chapter 15.

Chapter 9

Music analysis and music-analytic approaches to phrasing:
The third downbeat

9.1 Aspects of Music Analysis

As both specific and general music analytic approaches will be used in the following chapter, this section briefly introduces ideas of what analysis is and aims to do, and different specific methods of analysis that affect the following analyses to different degrees. This prepares for the interpretation of the results of this study through comparison of the analyses with the results of listeners' and performers' phrasing studies and identification of the musical features and phrase-parts identified for each piece.

The term 'musical analysis' refers to many diverse (sometimes mutually exclusive) activities. It is therefore difficult to define the boundaries and content of the field of 'music analysis'. However, according to Bent and Pople, underlying all aspects of analysis is the fundamental point of contact between mind and musical sound, namely musical perception and indeed some music perception research has been grounded in, or related to, music-analytic theories (Bent and Pople).

Music analysis is considered by some to be the resolution of musical structure into relatively simpler constituent elements, and the investigation of the functions of those elements within that structure. The relationship between the structures and elements proposed by analysis, and experiential, generative and documentary perspectives on music, has circumscribed analysis differently and has aroused debate (Bent and Pople). Analysis tends to use definable elements including phrase-units, harmonies, dynamic levels, measured time, bowings and tonguings (Bent and Drabkin 1987, p. 4).

Comparison of unit with unit, within a single work, between works, between works and abstract 'models' such as sonata form or a recognized style, is common to all kinds of musical analysis. The central analytical act is thus the test for identity and out of this arises the measurement of degree of difference or similarity. These two operations serve together to illuminate the three fundamental form-building processes: recurrence, contrast and variation (Bent and Drabkin 1987, p. 5). Analysis is based on the assumption that music 'makes sense' (Morgan 2003, p. 27).

9.1.1 Methods of, and approaches to, analysis

There is a wide variety of different methods of analysis including: Schenkerian analysis, Meyer's gestalt and later information-theory analysis based approach (1956), organic motivic analysis, 'functional analysis' (Keller 1956–7; 1957), feature analysis, formal analysis, semiotic analysis, style analysis. Some are related to one another, others are more distant. Some are mentioned here to give a flavour of the pre-occupations and musical features of some methods. For a more detailed discussion of the development of analysis as a discipline and the different theories, see, for example, Bent and Pople, and Bent and Drabkin (1987).

Schenker's unique view of musical composition is that tonal masterworks are 'projections' in time of a single element: the tonic triad. This projection comprises two processes: its transformation into a two-part *Ursatz* ('fundamental structure'), and *Auskomponsierung* ('composing-out'); the elaboration of the structure by one or more technique of prolongation (for example, Schenker 1906/1922/1935).

Meyer worked within the Gestalt concepts of *Prägnanz* and closure (chapter 1) and later with information theory viewing musical styles as culturally conditioned systems of expectations, and of musical meaning as deriving from the arousal, frustration and fulfilment of the expectations (Meyer 1956; Juslin and Sloboda 2001). Theorists including Schenker, Meyer (and Rothstein), assume musical structure to be goal-directed (section 9.2.1).

Réti's motivic analysis, by reduction of all the thematic material of a work to its common elements, produces 'cells' underlying its motivic material. These cells are non-rhythmic entities consisting of melodic contours of two or three intervals. Each can undergo transposition and inversion, and sequences of these can recur creating a 'thematic pattern' in and a symmetry or unity between movements in a work, which he considered a conscious act of composition. A movement can be set out, in non-rhythmic form with its melodic shapes grouped to reveal the motif forms, as a 'thematic song'. For Réti, the composer starts, not with a theoretical scheme, but with a motif which he allows to grow by constant transformation. Its growth is evolutionary. In time the composer makes a significant modification to the motif or picks up a detail from his elaborative material. Réti sees pieces as 'a musical improvisation … around a few motifs' (Bent and Pople; Réti 1951; Réti 1958).

9.1.2 The listening process and its representation in musicological analyses

These methods of analysis are score-based (usually excluding performance aspects). They may sometimes be so detailed that some aspects are not directly perceived by listeners. However, they may, at least in some ways, reflect how the music analyst 'hears' the music and describe and explain elements of how the music is structured.

Perception of phrase-structure and structures in general change while listening to a piece, depending on the new information. A criticism of some musicological analyses (and of some more explicitly psychologically based approaches such as that of Lerdahl and Jackendoff 1987) is that they provide just one 'interpretation' of the piece's structure. However, this does not mean that there is a necessary implication that all analyses are, or have to be, performed in this way. It is possible to reflect more closely the moment-by-moment changes gained through the addition of new information in the listening process, while at the same time, benefiting from music analytical approaches. How close discussions of music analysis can come to discussions of perception is illustrated by a discussion of unity and disunity in music (appendix 9).

9.2 Music-analytic approaches to phrasing

Phrasing is mentioned in passing in most music-theoretical discussions but has been analysed in detail by only a relatively small number of theorists.

9.2.1 Heinrich Christoph Koch

One of the earliest detailed analyses of phrase structure was provided by Koch in the second volume of his *Versuch einer Anleitung zur Composition*[79] (1787). This is a composition manual, a rule base, formalising contemporary conventions. Though it was intended mainly as a basis for composition, since it has also been used in theoretical discussions and even as a basis for theorists, notably Rothstein whose work is discussed subsequently, it is discussed here.

Though this is a composition *manual*, Koch explains that in general, only feeling can determine phrasing through the identification of the location and nature of

[79] Introductory essay on composition.

resting points in the melody (Koch 1787, 1983, pp. 3-4).[80] Lengths and punctuation formulas for these units do not offer a general characteristic trait for the location of resting points in melodies or for the completeness or incompleteness of sections (p. 4).

Indeed, Koch explains that anything related to speech or music must include more or less noticeable resting points, which break up the stream into larger and smaller sections: '[c]ertain more or less noticeable resting points are generally necessary in speech and thus also in the products of those fine arts which attain their goal through speech, namely poetry and rhetoric, if the subject that they present is to be comprehensible. Such resting points are just as necessary in melody if it is to affect our feelings. This is a fact which has never yet been called into question and therefore requires no further proof' (p. 1).

Through the more or less noticeable resting points, 'the products of these fine arts can be broken up into larger and smaller sections. … Just as in speech, the melody of a composition can be broken up into periods by means of analogous resting points, and these, again, into single phrases [*Sätze*] and melodic segments [*Theile*]' p.1. Koch tries to make the comparison with linguistic structure stronger and more direct, especially the identification of subject - predicate relationships in melodies, but abandons this as the two cannot be differentiated enough in melodic sections (pp. 4-6).

For Koch there are three defining features of musical periods; 1) the type of their ending (formulas/melodic punctuation) which let us clearly recognise more or less noticeable resting points, and 2) their length together with 3) a 'certain proportion or relation between them which can be found in the number of their measures once they are reduced to their essential components' (rhythm) (p. 1-2).

Koch describes the different types of sections (section 9.2.1.1), their lengths (section 9.2.1.2), the different types of endings (section 9.2.1.3) and extensions (section 9.2.1.4) and the ways in which phrases can be combined (section 9.2.1.5):

9.2.1.1 Section Types

Section types are primarily defined by the degree of closure at their end and can be:
1) An *incise* contains an incomplete idea.
2) A phrase [*Absatz*] may be complete or incomplete but cannot close.
3) A closing phrase [*Schlußsatz*] is a complete section identified by its closing formula (pp. 2-3). The *Absatz* cannot close a whole while the *Schlußsatz* can (p. 7).

Completeness in melodic phrases manifests itself in different ways:

---

[80] In the rest of section 9.2.1 for Koch (1787) the page numbers of the English translation (1983) of are given alone.

1) A basic phrase [*enger Satz*] may contain only as much as is absolutely necessary for it to be understood and felt as an independent section of the whole (p. 3). It is complete when it can be understood or felt as a self-sufficient section of the whole, without a preceding or succeeding incomplete segment fortuitously connected with it (pp. 6-7).
2) An extended phrase [*erweiterter Satz*] may also contain a clarification, a more complete definition of the feeling (p. 3).
3) A compound phrase [*zusammengeschobener Satz*] containing two or more phrases, complete in themselves, combined so that externally they appear in the form of a single phrase (p. 3).

In complete phrases, melody may cohere so closely that no noticeable resting points can be discovered which divide them up into incomplete segments. Or the phrases may contain these resting points and can be broken up into incomplete segments (p. 8) of which there may be more than two per phrase (p. 10).

9.2.1.2 Length

'Complete phrases need less or more length depending on whether they are basic or somewhat extended phrases' (p. 11). 'Most common, and also, on the whole, most useful and most pleasing for our feelings are those basic phrases which are completed in the fourth measure of simple meters. For that reason they are called *four-measure phrases* [*Vierer*]' (p. 11), four in simple meters or two in compound meter (p. 11). When four-bar phrases are broken down into smaller segments through resting points, the complete or incomplete incises that they contain are usually two bars long, dividing the phrase into two segments of equal length (p. 13). However, not 'all basic phrases are complete in the fourth measure; often such a phrase becomes complete only in the fifth or sixth, occasionally not until the seventh measure' (p. 14).

The basic phrases of more than four bars should not be confused with extended phrases of the same number; the former require a different treatment from the latter in the joining of phrases in a period (p. 14). 'If a phrase is complete in the fifth measure of a simple meter then it is called a *five-measure phrase* [*Fünfer*]' (p. 14) and can arise in three ways; 1) from a four-bar phrase through extension of two metrical units to two bars (p. 14), 2) by joining two unequal segments each of which is incomplete in itself and in which there is no extension, 3) when the motive in a bar of a four-bar phrase is continued in the subsequent bar (pp. 14-16). A six-bar basic phrase can also arise through (1) and (2) or by connecting a complete four-bar phrase with a preceding incomplete two-bar segment (pp. 17-18). For a seven-bar phrase, a five-bar phrase can be extended, or a complete four bar phrase can be preceded by an incomplete three-bar segment (pp. 18-19).

This description indicates that though Koch begins with the four-bar phrase, longer basic phrases of an even and odd number of bars are also possible. This emphasises that not all phrases are of the same length, or even of an even number of bars.

9.2.1.3 Punctuation signs/ending formulas

Koch then describes the different types of endings, beginning with the '*caesura*' (cutting): 'That place where a resting point is shown…where one section of the melody can be separated from the following one' (p. 19). This is the 'place' where the continuity of the melody is interrupted (p. 22), should always fall on the strong part of the bar (p. 19), and on a 'tone of an essential triad basic to that key in which the melody is rendered or to which it is turning' (p. 22). That tone can be unembellished or embellished with subsidiary notes. The various decorations of such a caesura note lead to different ending formulas of phrases and incises (p. 23), for example:
1) 'By means of striking afterwards [*Nachschlag*] other tones contained in the triad at its basis … In this case the caesura acquires an overhang [*Ueberhang*], or a feminine ending, which … can be mixed with passing notes and appoggiaturas in various ways' (p. 23-24).
2) By 'means of a suspension [*Vorhalt*] or an appoggiatura [*Vorschlag*] through which the caesura note is displaced form its proper position. Instead of it, a tone not belonging to the harmony comes on the downbeat, which necessitates the succession of the caesura note' (p. 29).
3) The 'space from a caesura note struck in the strong part of the measure or delayed by an appoggiatura is filled with notes until the tone with which the following phrase begins' (p. 34). The next phrase is here connected more closely with the preceding one (p. 34).

All this also applies to caesura of incises. However, as the incise is an incomplete segment of a phrase that must be followed by something more if a complete phrase is to result, the caesura of such an incise can be based on an underlying dissonant chord, usually on the fifth degree (p. 34-35): 'If the tone which makes the caesura of a phrase calls for the triad on the keynote… then the phrase itself is called a I-phrase [*Grundabsatz*]' (p. 36). If it requires the triad on the fifth degree, it is called V-phrase [*Quintabsatz*], often called a half cadence (p. 36).

'In phrase-endings, the root of that triad which lies at the basis of the caesura note is always placed in the bass' but there are also some cases where the third is used instead (p. 37). 'At the caesura of incises, the third of the triad in the bass is found nearly as often as the root itself' (p. 37). These can be with fermata (p. 35-36).

Three special notes belong to the cadence as the ending formula of the closing phrase: 1) the note of preparation of the cadence falls on the strong part of the bar. 2) The second note is the cadential note, which occurs on the weak part of the bar – it will imminently "fall" onto the keynote. 3) The closing tone or caesura note of the cadence is again on the strong part of the bar (p. 38). (1) and (2) are often decorated with subsidiary notes (p. 39). This ending formula is often augmented or diminished in length (p. 39).

9.2.1.4 Extended phrases

'A phrase is extended when it contains more than is absolutely necessary for its completeness' (p. 41). This can be done through:
1) Repetition of a segment of a phrase, in the same or other key (p. 41).
2) '[A]ddition of an explanation, an appendix, which further clarifies the phrase' (p. 45). This can be a section of the phrase, whose repetition makes the meaning of the phrase more emphatic, or may be an incomplete segment which is not yet present in the phrase but which is able to define its substance more closely. The extended phrase usually acquires two phrase-endings on the same root (p. 45). These additions are not included in the calculation of phrase length (pp. 47-8). When several cadences follow one another in a closing phrase the caesura note is usually not used in one of the last cadences, deceiving the ear in its expectation of the closing tone (*deceptive cadence*).
3) 'The last means of extending a phrase and defining more closely the feeling contained in it is parenthesis, or the insertion of unessential melodic ideas between the segments of a phrase' (p. 53). Phrases can also be shortened.

9.2.1.5 Compound phrases

The compound phrase (*zusammengeschoben*) is when '[t]wo or more phrases, of which each is complete in itself, may be connected with one another so that they either appear in the form of a single phrase or ought to be considered as only a single phrase within the period' (p. 54). The compound phrases can arise through various means: 1. '[S]tifling or suppression of a measure'; 'Two complete phrases in which the caesura note of the first and initial tone of the second phrase are one and the same degree of the scale are connected in such a way that the measure which contains the caesura of the first phrase is omitted, and the initial tone of the following phrase is at the same time considered as the omitted caesura note of the preceding phrase' (p. 55). 2. '[R]emoving that quality of the ending formula of the first phrase which characterizes it as a complete phrase' (p. 56). 3. An entangled phrase, in which the segments of two complete phrases are mixed so that a segment of the first is brought into the second and vice versa (p. 57).

9.2.1.6 Koch: Summary

For Koch, phrases are primarily defined by their cadences and especially the degree of closure provided by the cadences. These can be of different relative lengths; the incomplete sections, when put together with their completion, become longer, complete phrases. Koch seems to refer two such levels as phrases. He distinguishes between basic phrases (which can vary in length) and extended or compound phrases, which are considered to last as long as the basic phrases of which they are elaborations. Despite the different names for different length phrases, the overriding methods of distinction between phrase types for Koch seem be cadences and voicing of the final chords. Koch mentions fermatas as accentuating the caesuras, but not as defining them. He mentions repetition but only as part of expansion of phrases. Koch does not mention pitch intervals, lengths of individual notes, or changes in motive or texture or any of the other musical elements mentioned by some other theorists (chapter 8). Rothstein, who based much of his theory on that of Koch, also concentrates primarily on tonal characteristics of phrases.

9.2.2 Rothstein

One of the most in-depth theoretical descriptions and discussions of the phrase is given by Rothstein in *Phrase Rhythm in Tonal Music*. His theory is based on Koch's and Schenker's ideas. Here parts of his work are discussed. For Rothstein, the history of phrase rhythm in tonal music is closely bound up with the history of the influence of dance and folk music on art music during the tonal era (Rothstein 1989, p. 34).[81]

9.2.2.1 Tonal Motion

The overriding aspect of the phrase is tonal motion: 'a phrase should be understood as, among other things, a directed motion in time from one tonal entity to another; these entities may be harmonies, melodic tones (in any voice or voices), or some combination of the two. *If there is no tonal motion, there is no phrase*' (p. 5).

For Rothstein, the 'goal of a phrase, the cadence, is not by its nature also a new beginning. Phrase endings *may* be used as new beginnings – this is what happens in the case of the phrase overlap (…) – but a *double entendre* of this sort is not inherent in the nature of the phrase. By contrast, except at the very highest levels metrical patterns are perceived as moving toward downbeats and thus toward new beginnings. (In general, only when a piece ends with a metrical downbeat is that downbeat not also a new beginning, and even then the continuation of a new metrical unit may be implied.) Thus a continuous renewal of rhythmic energy seems to be inherent in the nature of meter to a degree that cannot be said of phrase structure. With phrase structure it is the *tonal* relationships, especially the feelings of harmonic completeness or incompleteness that tend to provide impetus for continued motion' (p. 28).

9.2.2.2 Rhythm

Rothstein arrives at the rhythmic dimension of phrasing from the tonal one; when a cadence (the end of a tonal motion) occurs with the end of rhythmic segment, the two feelings of completion reinforce each other. The feeling of constant

[81] In the rest of section 9.2.2 this Rothstein 1989 is referred to with page numbers only.

movement toward a goal is intensified when the harmonic and melodic goal is underlined by rhythmic means. 'A phrase is first of all a unit of tonal motion. However, since every motion in music must take place in time, the phrase is simultaneously a rhythmic unit.' The rhythmic aspect arises principally from its length; 'the length of successive phrases that gives rise to phrase rhythm in the most literal sense' (pp. 27-8). For Rothstein, phrase rhythm describes rhythmic phenomena involving phrases and hyper-bars (p. 12). He explores rhythmic expansion in tonal music; the process by which a relatively small and regular rhythmic unit is transformed into a larger and less regular one.

9.2.2.3 Hierarchy

For Rothstein there can be more than one level of phrasing: 'Large phrases often contain smaller ones (and may themselves be contained by still larger ones); small phrases may contain subphrases' (p. 11). Both metrical and phrasing structures can be organised hierarchically, however, though they 'may be analogous, they are not equivalent' (p. 11). Hypermeter is the combination of bars on a metrical basis, including the recurrence of groups of an equal number of bars and a definite pattern of alternation between strong and weak bars. Phrase structure, on the other hand, is the coherence of musical passages on the basis of their total musical content – melodic, harmonic and rhythmic – and can be determined with the help of careful melodic and harmonic analysis. Other elements, such as articulation and dynamics are generally supportive rather than determinative of phrase structure, at least in tonal music. The best available means for the melodic and harmonic analysis is the Schenkerian method, as it reveals precisely the underlying tonal motions. Hypermeter and phrase structure may coincide or they may not; their agreement or conflict represents a basic compositional resource (pp. 12-13). The tension between phrase motion and hypermeter is increased when the phrase does not begin on a hypermetrical downbeat i.e. when the two are out of phase (p. 29).

Subphrases, according to Rothstein are tonally incomplete (p. 30). Motivic repetition is one of several factors that may determine subphrase grouping, but an especially powerful one (p. 30). It is relatively uncommon for a subphrase to coincide exactly with a single bar (except for example, Chopin's Prelude in C minor Op. 28, No. 20) (p. 31).

In general, divisions between subphrases tend to be subtler and more ambiguous than those between complete phrases. For musical analysis, it is sometimes pointless to insist on a single set of subphrase divisions for a given phrase. A certain ambiguity in this respect may be the source of a melody's fluidity (p. 32).

9.2.2.4 Legato

Rothstein also points out the difference between legato and phrasing symbols:

> 'Phrasing (noun) 1. The delineation and internal shaping of *phrases* … by a musical performer. Includes both the joining of notes into phrases and the separation of these phrases from each other.
> 2. The *legato* performance of notes under a slur; or, the *legato* performance of notes as if they are under a slur. Involves the physical connection of notes (*legato* playing) regardless of their position within the *phrases*.
>
> The 'joining of notes' in the first definition is a connection according to musical meaning – that is, according to tonal (and perhaps rhythmic) motion; while the "connection of notes" in the second definition refers only to legato articulation, regardless of its purpose.' The problem comes when we use slurs (symbol for legato articulation) to indicate phrasing.
> (Rothstein 1989, p. 11)

'If the meaning of "phrasing" could be restricted to the first definition given above, it would prove a valuable term indeed. It would then refer to and summarize the entire panoply of means – dynamic, rhythmic, and articulative – by which a good performer communicates the phrase structure of a piece of music' (p. 12).

'All tonal music, with the arguable exception of the Prelude to Wagner's Das Rheingold, is composed of phrases in the sense used here. Not all tonal music is composed of hypermeasures, however' (p. 13). Improvisation-like Baroque toccatas, Classical fantasias and recitative-like passages for example, often avoid them. Hypermeter is most strongly evident in pieces intended for or suggest dancing (p. 13).

Having described the basic characteristics of the phrase and distinguished it from other musical elements, Rothstein discusses the relationships between phrases.

9.2.2.5 Antecedent and Consequent Phrases, Periods

Antecedent and consequent phrases together form a larger phrase or a period. A period is a large phrase that contains smaller ones. So 'phrase' is a generic term while 'period' is a species of phrase (p. 20). The antecedent ends with a half cadence, and the consequent begins like the antecedent but leads to a full cadence. The close of the half cadence is weaker than the full, but is sufficient to create an impression of a minimally complete thought and considerable sections often end with them (p. 17).

The antecedent usually involves an interruption of 3-2||3-2-1 or 5-4-3-2||5-4-3-2-1. Thus melodic tension is added to the harmonic tension of the half cadence. Both are resolved by the full cadence at the end of the parallel period. An exception is when the second of the two phrases in a parallel period ends in a different key. When there are two phrases that together form a period but that are not antecedent - consequent in relation, they are called a fore-phrase and after-phrase (p. 18).

9.2.2.6 Phrase Overlap

Two phrases (or subphrases) overlap when the last note (or chord) of the first phrase acts simultaneously as the first note (or chord) of the second. More rarely, more than one note or chord may be common (p. 44).

Rothstein refers to Lerdahl and Jackendoff (1987), who use the term 'elision' rather than 'overlap' to denote cases in which the ending and beginning chords of the two phrases would not be identical if the phrases were separated (chapter 8.1). This is a logical distinction, since one hypothetical chord (usually the last chord of the first phrase) has indeed been omitted or elided in the conjunction of the two phrases. But since the perceived effect is still one of phrasing overlapping – the listener 'hears' the elided chord subsumed in the actual one – Rothstein uses the term overlap whether or not an elision is present (p. 46). A phrase overlap is most likely to occur when the first of the two phrases ends either at (or just after) a hypermetrical downbeat (p. 48).

'[C]omposers use phrase and subphrase overlap chiefly to secure a greater feeling of continuity in the melody. When two melodic segments overlap, a point of melodic punctuation - to use an 18th-century expression - is eliminated or transcended. Overlap is particularly frequent in those styles in which surface melodic continuity is most highly prized' especially those of the Late Baroque and Romantic periods (p. 51).

9.2.2.7 Internal characteristics

Rothstein discusses the internal characteristics of the phrase, particularly in the context of a 'drive to the cadence', the 'characteristic of many phrases as they approach their goals' (p. 22). 'Both melody, and especially, the bass begin moving in much faster note values than before…It is especially frequent in consequents and after-phrases, where the cadence concludes not only a small phrase but a larger period' (p. 22). Pre-cadential acceleration serves several purposes; change in harmonic rhythm helps to signal the coming cadence, often in conjunction with other changes (e.g. texture or surface rhythm) and provides a climax of rhythmic activity just before the cadential relaxation. Increased harmonic excitement counteracts the inherent harmonic predictability of the pre-cadential situation, holding the listener's interest to the phrase end (p. 22-3).

9.2.2.8 Lead-ins

A lead-in is a melodic unit that is less complete (and usually shorter) than the phrases that it connects and entails overlaps between it and the beginning of the following phrase. It is not a subphrase because it is not part of any complete phrase but merely a link between two such phrases (p. 51).

At larger levels of phrase structure, it is possible that a complete phrase may act as a lead-in connecting two larger periods. For example, at the end of the development section in a sonata form, a period generally ends with a dominant of the main key (half cadence). The next period begins the recapitulation, usually starting with the tonic harmony. A phrase leading from dominant to tonic may connect these two periods. This would be a lead-in at a larger level (p. 52).

Metrical reinterpretation seems to occur in conjunction with an overlap in the phrase (or subphrase) structure. Apparently without an overlap there would be no reason for the listener to assume a reinterpretation in metrical structure. Overlap refers to phrase structure, reinterpretation to metrical structure. Overlap often occurs without reinterpretation, but reinterpretation never occurs without overlap (p. 52).

9.2.2.9 Elongated Upbeats

Elongated upbeats precede the first bar of a hyper-bar and lasts at least a full bar. Generally, an upbeat is 'elongated' only if its inclusion results in the appearance of one or more extra bars between hyper-bars, or preceding the first hyper-bar of a piece. These extra bars are not counted as part of any hyper-bar, but may force adjustments to our usual sense of hypermeter (p. 39). Metrically, therefore, an elongated upbeat disrupts the hypermeter by delaying the beginning of a hyper-bar by a bar or more. When the elongated upbeat occurs between two hyper-bars, it separates them, and thus has the effect of suspending the hypermeter momentarily (p. 56).

It is more difficult to make any general statement concerning the effect of an elongated upbeat on phrase structure. The fact that an extra bar is heard as an upbeat implies that it is an upbeat to something – that it groups with whatever follows it. This is certainly true in the vast majority of instances. It is also possible, however, to embed an elongated upbeat between two phrases in such a way that it relates to both the preceding phrase and the following one. Schenker (1979) cites an example from Beethoven's Seventh Symphony (bb. 61ff) that includes two elongated upbeats (p. 57).

9.2.2.10 Successive Downbeats

Rothstein also discusses successive downbeats. For example, in a melody and accompaniment texture, a hypermetrical downbeat may be 'split', appearing first in the accompaniment, then in the melody (p. 58). The accompanying component may begin a phrase one or two bars before the melody, playing (sometimes repeatedly) an accompanimental figure. If this figure sounds a single harmony and the melody then enters over the same harmony, the resulting effect is often that of a single hypermetrical downbeat split between the two parts of the texture (p. 63).

9.2.2.11 Regular and irregular phrase construction

For Rothstein, regular phrases are of duple construction and irregular are of non-duple. Duple phrase lengths are the 'best' or 'most natural' while others may be used to good effect. Some of these non-duple phrases may be produced by modifying regular (i.e. duple) phrases in various ways; others, however, cannot be so produced and must be considered as irregular phrases independent of duple models (p. 33).

According to Schenker, preference for duple organisation is innate to humans for physiological and psychological reasons. This leads to a powerful normative influence to duple structures, which is one reason why many non-duple phrases can be understood as modified duple ones. However, non-duple structures do not always depend on duple models (pp. 33-4). For example, phrases can be constructed on the basis of a number of principal tones (Schenker 1979, pp. 120, 127, Figures, 40, 7, and 148, 6; Rothstein 1981, pp. 70-72; Schachter 1987). Schachter cites Schubert's 'Wanderers Nachtlied' D. 768 (b flat major) where a phrase of five half-bars corresponds to the five principal tones in the melody (chapter 2.4.2).

It is also possible to 'sneak them in … under the cover of a larger duple phrase (or period) that contains them' (p. 35). When a large duple phrase is divided into non-duple segments, the hypermeter is likely to be suspended or even nonexistent. Hypermeter is strongly influenced by phrase structure, and the perception of duple hypermeter cannot easily survive a strongly asymmetrical phrase construction. If the two- and four-bar levels of hypermeter are not present, it is difficult for an eight-bar hypermeter to emerge convincingly.

9.2.2.12 Expansions

Expansion describes rhythmic operations that may transform a phrase into a larger one and is a kind of embellishment (p. 64). It surpasses some contextually established norm of phrase rhythm. There may be cases in which the sensitive listener hears an expansion, but no prototype or basic phrase is readily apparent (p. 93). The references from which an expansion departs are generally metric and tonal. Most basic phrases have a well-defined hypermeter, which usually matches the prevailing one (if any). Like an elongated upbeat or cadenza-like insertion, the expansion temporarily suspends the hypermeter without breaking it. Often, the underlying hypermeter can be heard without much difficulty. At other times, greater effort is required (p. 65).

Some types of expansion do not preserve the tonal skeleton (middleground) of the basic phrase. According to the definition of phrase expansion, sequentially constructed phrases are only regarded as expanded if the sequence is not essential to the middleground structure of the phrase (p. 94).

Once a phrase is perceived as complete, its rhythmic completeness cannot be revoked by the addition of new material following the cadence. Phrases may be rendered *tonally* incomplete, in retrospect, by changing the subsequent tonal direction (leading to a different cadence). The converse is represented by phrases that do not fill all of a predetermined temporal unit, and to which a suffix is added partly in order to fill out the requisite length (p. 96).

An expansion of the basic phrase transforms part of the original hypermeter, stretching it perhaps to several times its original length. The original hypermeter is still, in principle, said to exist. But expansions tend to fall into their own hypermetrical patterns, resulting in conflict between the surface hypermeter in the expansion and the underlying hypermeter of the basic phrase. Often, it is possible for the listener to perceive both hypermeters simultaneously; at other times, the underlying hypermeter may be pushed so far into the background that it virtually disappears (p. 97).

Phrase expansions often create complex, multilayered hypermetrical structures that consist of a surface hypermeter and one or more levels of underlying hypermeter (the main one being that of the basic phrase). The perceptibility of an underlying hypermeter may vary: it may be nearly self-evident, vaguely intuited, perceived only with difficulty, or missed altogether (p. 99). Examples include:

External Expansions: Small prefixes often take the form of an accompanimental figure, preparing for a melodic entrance (p. 68). The underlying phrase structure does not include the prefix (p. 70). It is usually clear where the prefix ends and the basic phrase begins: there is either a change of harmony (often V-I), or melodic pattern, or an overlap between the prefix and the basic phrase.

Suffixes are the most common type of expansion. The location of its start is not as clear as for the prefix. Suffixes usually do not begin with a change of harmony; they extend the basic phrase's closing harmony. Small suffixes are common after full and half cadences (signalling increased tension and expectancy, used to prepare important thematic entrances). Large suffixes are also common. Codas are usually large suffixes, while codettas may be either large or small. Large suffixes may also occur after a full cadence (rarely after half cadences). Large suffixes are especially common in sonata forms: the 'closing theme' (of exposition and recapitulation) is typically a large suffix to the second group (p. 71). Several suffixes may occur consecutively, so that the suffixes have suffixes. The essential quality of a suffix, as Riemann states, is the extension of a goal already reached.

Internal Expansions: These include repetitions within the phrase (p. 74) and delay of the authentic cadence by means of a deceptive one.[82] This is not a suffix, because the true phrase end arrives only with the authentic cadence.[83] If an expansion is present, the length of the phrase up to the deceptive cadence should be roughly equal to or greater than that of the surrounding phrases and it should be possible to substitute an authentic cadence for the V-VI motion thus obtaining a basic phrase (p. 78).

Expansion by composed-out Deceleration or Fermata: Composed-out deceleration occurs typically just before the phrase's cadence. It frequently affects not the surface rhythms but some more abstract level of motion such as harmonic rhythm or even the rhythm of a middleground progression (p. 83).

Expansion by Parenthetical Insertion: For Koch 'Parentheses' is 'the insertion of unessential melodic ideas between the segments of a phrase' (p. 87). In Classical works, where they are most common, this plays a variety of roles, including: 'asides', insertions just before the final cadence, and interruption of concluding phrases (p. 88).

9.2.2.13 Form

For Rothstein, a discussion of phrase rhythm must become a discussion of form. Hypermeter plays at best a secondary role in the creation of form because of the nature of meter. Metrical schemes require repetition of units that are, at least conceptually, of equal size. As the size of the units increases repetition becomes increasingly problematic. A whole composition is not generally repeated (except in strophic songs) and therefore cannot be considered a single metrical unit. A piece however can be a single phrase as phrases do not require repetition.

The larger sections of a composition are delineated not by their metrical structures (which cannot be a primary guide to musical form) but by their cadences, which are generally obvious (pp. 102-3). For Rothstein there are two levels of form: Outer form - the thematic aspect of a piece and its layout into phrase and periods, and Inner form - the tonal dynamic of a work –its large-scale harmonic and linear layout (p. 104).

---

[82] Occasionally, a deceptive cadence is not followed by the expected authentic cadence though no general statement can be made about such cases (p. 80).

[83] The melodic line in this expansion may repeat part of the basic phrase or may deviate from it. The bass, however, normally returns to the dominant in order to cadence V-I; often a part of the bass line is actually repeated (p. 78). Koch also described deceptive cadences in which the melody rather than the bass is the agent of deception. For example, where an imperfect authentic cadence on $\hat{3}$ substitutes for an expected perfect cadence on $\hat{1}$. Such deceptive cadences, for Rothstein, are parenthetical interpolations (p. 80).

9.2.2.14 Rothstein: Summary

Rothstein, basing his work on that of Koch, prioritises complete tonal motion as the main defining feature of phrases and relates this to the phrase's rhythmic structure, which is different to, but interacts with, hypermetrical structure. Having established the basic structure, he explores types of phrase extension and relation between phrases.

9.2.3 Music-analytic approaches to phrasing: Summary

Koch and Rothstein define phrases primarily harmonically: For Koch, they are defined by their cadences and especially the degree of closure provided by cadences, for Rothstein, they are tonal and rhythmic goal-directed entities. Both, especially Rothstein, concentrate on tonal motion within the phrase. In this way, both concentrate on the phrase end and the arrival at the end more than other phrase parts. They distinguish between different types of phrase endings, broadly splitting them between closed and open phrases (requiring continuation), and identify several types of shortening, extension and combination of phrases.

Both discuss a basic structure that is modified by elision or expansion of different types. There is often an identification of the 'basic' phrase structure which can be distinguished from its modifications. Both discuss different types of phrases and place these in the context of periods. They discuss both hierarchical relationships and the relationship between adjacent phrases (such as antecedent-consequent pairings). Both discuss the lengths of phrases (in terms of bars), and highlight the preference for four-bar phrases, but do not define phrases purely through their lengths. Both explore the relationship between phrasing and metrical structure, clarifying that they are different but do interact.

Neither mentions pitch intervals, lengths of individual notes, or changes in motive or texture or any of the other musical elements mentioned by some other theorists (chapter 8). Elements such as fermatas and phrase lengths are discussed but are not central or necessary for the identification of phrases. The elements that they do discuss, along with those discussed in the comparisons chapter, inform the analysis of the case-study pieces (chapter 10).

Chapter 10

Musicological analyses of the case-study pieces and comparison with the results of the current phrasing studies:
The fourth downbeat

10.1 General Introduction

The previous chapters have shown that there are systematic responses to the tasks of the identification of phrase starts (PS), phrase ends (PE) and the beginning of the expectation of the end (EOP) by listeners and by musicians in their written responses, and that performance characteristics of dynamics and tempo change can also be systematically related to phrases. The next stage is to explore 1) whether there is a systematic relation between responses and musical features of the pieces and what it might be, and 2) the nature of phrases, how they are constructed, and how they relate to one another.

In this chapter, phrase identifications of the same pieces by music-analysts and more general music theoretical approaches are discussed. Musicological analysis, being based primarily on the analysts' 'hearing' of a piece, provides not only another view of phrase perception but also the tools and terminology to systematically explore the possible musical reasons for these systematic responses.

The following sections give analyses of each one of the case-study pieces using various analytical approaches including both published and new analyses. The published ones do not always take phrasing as their central topic, but often use the idea or refer to musical elements that may affect phrase perception. Here, each analysis section includes suggestions of the phrase-structure of the piece as derived from the musicological analyses. These are used in the process of musical feature identification for each piece. The features are compared with results of the previous chapters.

This leads to the identification of: 1) possible musical features and their characteristics that contribute to phrase structure and its perception, and 2) a more detailed description of what phrases are, what 'parts' they are made up of and to what extent they are identified by listeners, performers and analysts (the methods and general results for which were presented in chapters 2-9).

Through 1) the analyses of the positions identified by participants as phrase parts and positions interpreted as indicating phrase parts in performance contours in the case-study pieces and the musical features that occur in these positions, 2) the musical analyses of the case-study pieces, and 3) more general observations through general music theory and theories of music perception and the explorations of specific methods (chapters 1 and 8), the following features are identified and investigated in the following analyses of the case-study pieces.

Features

Musical Features
Bar-line
Four-bar template (and four-bar template from start of melody)
Position in bar of opening
Pitch Jump (relatively large pitch intervals)
Long note
Rest
Change in Motive
Change in Texture
Following Cadential Progression
Following Voice-Leading Progression
Exact Repeat
Inexact Repeat
Imitation in Lower Part

Performance features
Change in Tempo
Change in Intensity

For each piece (in each section), these are plotted on the timeline. The graphs show the coincidence of several musical characteristics in certain areas. There are some areas where the coincidence of features is greater than in others. A comparison of the listeners' responses and performance contours with aspects identified in the analysis, particularly the musical features follows. To enable direct comparison, all the graphs relating to each piece can be found together in appendix 3.6. The features are also gradually categorised according several aspects: their area of presence, impact (instantaneous or over an area) and 'function' or result (predictive, instantaneous and retrospective). These are discussed in more detail in chapters 11-13.

10.2 Swimming in Bach: Wiewohl mein Herz in Tränen schwimmt from *St. Matthew Passion*, by J.S. Bach (Bach Passion)

10.2.1 Introduction

Bach's music is known for its counterpoint and its independence of parts, the combination of which results in complex harmonic relationships. In 'Wiewohl mein Herz in Tränen schwimmt' (No. 12 in Bach's *St. Matthew Passion*, 1727, Bach Passion) each part has its own complexities, which are in some ways compounded rather than alleviated by the combination with other parts.

This movement is a setting for soprano, two cor anglais and continuo of a meditative 8-line text by Christian Friedrich Henrici (pseudonym Picander). This recitative, like others in the *St. Matthew Passion*, is the middle movement of a three-movement group: 1) Biblical narrative (No. 11). 2) Reflective comment on the biblical narrative in a recitative (No. 12). 3) Prayer in an Aria (No. 13).

Vocal music has characteristics not present in purely instrumental music including the relation between text and music (chapters 2 and 15). In chapter 2 it was seen that: 1) when the original text has a clear structure, the musical phrases often follow its poetic (couplet) structure and 2) when there were contradictions between the poetic structure of the strophe and the syntactic structure of the sentences, these were reflected in musicians' phrase identification. The musical structure of recitatives[84] is usually regarded as having a musical structure that is even more closely related to that of the text. It usually remains simple, the soloist's

[84] derived from recitare, to recite

melody following the text's rhythm and structure, and held chords in the accompaniment that only emphasise the cadences.

However, in this recitative, the orchestral part is much more active, and although each text line is separated from the next by a rest, this alone does not coincide with a clear phrase structure. Bach's recitatives, though perfectly geared to the German language, are basically an adaptation of the idioms of Italian opera and there is no difference in style between his use of it in his sacred works and his secular ones.

10.2.2 The piece

10.2.2.1 The text and its musical setting

Several characteristics of the text suggest various possibilities for its subdivision, some of which are suggested here (the text is given in appendix 3.6.1).

Lines 1-5 and 6-8

The text has the following rhyme scheme: aa bb c dd c, indicating a possible division into two parts aa bb c and dd c. This division is reflected in some musical characteristics. For example, the rest in the vocal part between the end of the fourth line and the beginning of the fifth is the longest in the piece. It is also the location of the 'cadence' that is most like a 'normal' perfect cadence (except the end). Table 10.2.2.1 shows a summary of some text structure and musical characteristics.

Table 10.2.2.1: Summary of some text and musical characteristics

Text line no.	No. of syllables	Metrical stress in text	Rhyme-scheme	No. of beats	Line ends on
1	8	u/u/u/u/	-a	6	Half bar
2	8	u/u/u/u/	-a	6	Down beat
3	10	u/u/u/u/u/	-b	6	Half bar
4	8	u/u/u/u/	-b	7	Weak beat
5	9	u/u/u/u/u	-c	5	2nd quaver of half bar, longest rest, nearest to 'normal' cadence
6	1	u/u/u/u/u/u	-d	6	2nd quaver of half bar
7	7	u/u/u/u	-d	4	2nd quaver of half bar
8	9	u/u/u/u/u	-c	6	2nd quaver of down beat

Lines 1-2, lines 3-8

Concurrent to the above line division there is, broadly speaking, a change of tone in the text from mourning in lines 1 and 2 to hope in the rest (appendix 3.6.1). A division into these two sections seems to be reflected in a number of musical features. For example:

1) Thematically and texturally, the first 3 bars (setting lines 1 and 2) have continuous triplet semi-quavers that end with the last note of the second line. After this, the theme and texture change as the semi-quaver-triplets are separated by held notes.
2) Metrically, the first line finishes on the half-bar and the second on the down-beat. This is the only place in the movement where the voice finishes on the strong part of the beat on the first beat of the bar (see table 10.2.2.1 above). The next line ends again on the weaker beat, so we could expect the next line (the second half of the couplet) to end on a strong beat again. Instead, however, the line ends on the second beat of the bar. From here to the end, the lines all end on the second quaver of the beat (following the stress pattern of the text) and it is only in the accompaniment that there is an end at the beginning of a beat, but even this, is on a half bar. The first couplet, therefore, is the only one that is metrically structured as opening and closing.
3) In terms of voice-leading, bars 1-4 (setting text lines 1-2) outline a 'closed' (ascent-) descent of (f#^{2})-a-g#-f#.
4) The change from lines 1-2 to the following couplet is reflected harmonically (section 10.2.2.2).

The second part (lines 3-8) is also divided thematically: The first two lines mourn in tears, and the next three celebrate Jesus' legacy, the next two describe his life on earth, and the final mentions the end forming a 1-2, 3-5, 6-8 line structure. The musical characteristics coinciding with this structure are discussed in section 10.2.2.2.

Stress Patterns

The stress patterns of each text line seem to be reflected in the metrical position of the start and end of each line; all the lines begin on an unstressed syllable, and all start on an upbeat to the first or third beats of the bar. However, the number of syllables in each text line does not remain the same in the setting. The vocal line mainly has quavers, and usually one quaver per syllable. However, in some cases 'key-words' of the text are accentuated, with these (besides those that coincide with the end of a line) having the longest (and often highest) notes: Tränen, Jesus, Fleisch, Blut, Kostbarkeit, and liebt. This seems to be reflected in performance characteristics, in the use of both dynamics and of tempo (section 10.2.4). The rhythmic structure of the setting, then, does not seem to be the same as the original poem.

10.2.2.2 Harmony and the text

The preceding movement (No. 11) ends on its dominant (G). The following one, (No. 13) begins in GM and cadences at its end on b minor. This movement (No. 12) overall seems to fit more closely with the key of the Passion up to No. 58^d of E minor.

Indeed, the first harmony is a diminished chord on D#, implying E. However, after a bar and a half, the change of harmony is to EM^7 with the 7^{th} in the bass. This arrives with the last word of the first text line 'schwimmt'; we really are left 'swimming'. The harmony change is half a bar later (bar 3) onto an $F^{\#}m^{(7)}$ chord which could lead to G but instead is followed by E^c on 3.5. So the first consonant harmony of the piece is not a resolution; if anything it indicates a beginning of a Ic-V-I progression in E.

Instead, however, $F^{\#}M^7$ arrives coinciding with the last note of the second text line. $F^{\#}M^7$ is the dominant of B and, for the first time in the movement, an expectation is confirmed by a B minor chord in bar 5 which coincides with the last word of the third text line. This coincides with the first change of mood in the text. The first couplet expresses deep sadness. The next line begins the new, more positive subject and it is on the last syllable of this line (the 'freut' of 'erfreut') that the B minor resolution arrives (see appendix 3.6.1 for the text).

Clarity of expectation and resolution does not last and the diminished harmonies return (on $G^{\#}$, $C^{\#}$, $D^{\#}$) finally reaching E^b minor on bar 8, leading to Ic-V-I in E minor in bars 8^3-9. Again, there is a change of subject (to Jesus' life on earth). Here, the end of the text (and vocal) line does not coincide with the resolution of the cadence but occurs half a bar earlier. This happens again at the movement's end. In bar 10 there is a relatively clear move of B dim to C Major (coinciding with the last word of the sixth text line) followed, in bar 11, by F# dim and G^7 leading to Ib-IV^7-Ic-V-I in C major (possibly relating to the opening harmony of the previous movement).

There seem to be local anchor points: bars 5 (Bm), 9 (Em) (10, CM), 13 (CM). In bars 5 (and 10) the melody line (and the text) coincides with the harmony of the accompaniment. However, the melody ends on $\hat{3}$ of B (bar 5) and C (bar 10), not on $\hat{1}$. The B of bar 5 and CM of bar 10 both follow sections in which both the cor anglais have temporarily stopped their otherwise almost continuous triplets. In bar 5 the B is preceded by a semibreve tied to a crotchet followed by a crotchet rest. In bar 10 the C is preceded by a dotted minim and a crotchet rest. The long notes and rests do not coincide with the resolution of the cadence; both occur before it. The temporary hiatus in bar 4 can be seen as related to the change in subject of the text (mentioned above). The triplets end with the end of the couplet about 'tears'.

The cor anglais restart on the B minor of bar 5.5 while the vocal line has a weak arrival. This may be because of the arrival on $\hat{3}$ instead of $\hat{1}$ mentioned above. It may also be because the previous end of a textual line (bar 4) was stronger than this one; from the start of the movement, the vocal line begins on the high f# and leaps down to the a and the 'melody' continues until the G# to the end of the textual line. The next text-line ends on f# (the root of the chord). This is the starting note of the vocal line and falls on the first beat of a bar. This provides a reason to interpret the structure as being a weak close followed by a strong close and the same pattern is expected again in the following bars (bars 4.25 - 7). In some ways then, despite the tonal resolution, a strong close should not be expected on bar 5.5.

This is further supported by the text's rhyme scheme of a a b b c for the first 5 lines. Although the next (fourth) line may be expected to coincide with a strong close (bar 7), this does not occur. Instead, the line ends on the second beat of the bar with a diminished harmony on C#. Instead, the strong close is delayed to the end of the next line (bar 8). This coincides with the subject matter of the text: The first two lines mourn in tears, and the next three celebrate his legacy (the next two describe his life on earth and the final indicates the end). However, the ending on bar 4, despite the strong features of returning to the opening note, coinciding with a long note in the cor anglais and the vocal line sounding the root of the chord, this is a 7th chord whose root (F#) is only arrived at in the bass at the beginning of bar 5. Here it becomes evident that it is acting as a dominant of B, which is fulfilled on 5.5. The ending of the both vocal line and the text is therefore stronger at the beginning of bar 4 (supported by the long note in the cor anglais) while that of the harmony only arrives half way through bar 5.

Unlike in bars 5 and 10, in bars 8-9 and 12-13 the vocal line and orchestral parts do not coincide. In bars 8-9 the solo ends first (on e) while the EM harmony is only reached in the orchestral part on bar 9. In bars 12-13, the melody ends on the first beat of the bar while the orchestral part ends on the half bar. In bars 8 and 13 the melody anticipates the root of the cadence resolution that then arrives in the orchestral part. In both cases, the root in the vocal part is sounded against I^{c} in the orchestral part i.e. the beginning of the cadential progression. Within the vocal line, the dominant note of each cadence is highlighted on the first beat of the bar; on bar 8 with b and on bar 12 with g^{1} on *liebt*. The latter is more prominent being the highest note since the g^{1} on *Jesus* of bar 3. The positions of the clearest cadences are summarised in relation to text structure and other musical features in table 10.2.2 below.

10.2.2.3 Melody

Being a recitative, this melody has a different style to songs or arias. However, it also differs from other recitatives, which often have restricted or even almost monotone vocal lines. Here there are large intervals from the start and there is little regularity.

Unlike in the vocal music studied in the introductory study (chapter 2) there is no lengthy repetition in the melody, even where it may have been expected (such as from bar 9.375). However, there is a figure used at the closes of text lines of bars 8, 10 and 12-13, and is underlies the vocal line in bars 3-4. From close to close, the upper line outlines a descent from the opening f# to c. Bars 8, 10 and 12-13 coincide with text endings but only 10 coincides with harmonic resolution. The location of these figures in comparison with the location of other musical features and the text lines is given below in table 10.2.2.

10.2.2.4 Metrical Structure

The vocal line always starts half-way through a crotchet beat but can begin on the upbeat to a bar (bars 2.875-3, 5.875-6, 10.875-11 and 11.875-12), to a half bar (bars 1.375, 4.375 and 9.375) or half way through the third beat (bar 7.625). It can end on the third (bars 2, 5, 8, 10 and 11) first (bars 4 and 13) or second (bar 7) beats, while the orchestral part tends to start on down beats (usually bars or half-bars). Much like in the Bach Suite (chapter 10.3), it seems that the half bar and bar have equal importance. The position of the main beat is constantly changing and is even more confusing because of the mismatch between voice and accompaniment.

10.2.2.5 Texture

Changes in texture occur when the cor-anglais' triplet-semiquavers change to longer notes or rests but rarely coincide exactly with vocal starts or ends (as shown in table 10.2.2 below).

Table 10.2.2 summarises the harmonic and melodic cadential figures and changes in texture in relation to the text line positions

Table 10.2.2: Summary of text structure and indicative musical characteristics (in bars)

Text Lines	Clear Cadences	Melodic Cadential Figures	Changes in Texture
1 – 2.5	1 – 5.5	1 – 4	4
2.75 – 4			
4.25 – 5.5		4.25 – 8	
5.75 – 7	5.5 – 9		5.5, 6
7.5 – 8.5			7, 8.5
9.5 – 10.5	9 – 10.5	9.5 – 10.5	9.5, 10.5
10.75 – 11.5		10.75 – 13	11, 11.5
11.75 – 13	– 13		12.5, 13.25, 13.5

10.2.3 Listeners' Responses

10.2.3.1 Areas of high listeners' response

The areas are discussed according to the high-response areas of the listening studies (graphs 3.6.1.1-5, appendix 3.6.1, and chapters 3 and 4).

Bar 1

The voice begins after the orchestra. The responses indicate that listeners assume that the first PS is start of the piece and do not 'interrupt' it with the start of the vocal line. The tendency to 'interrupt' does increase with increased listenings (or familiarity).

Bar 2

MIDI responses in the second half of the bar may be related to the harmonic and voice-leading expectations and the long note and dotted rhythm leading to the g# in the vocal line. The vocal line has a rest on the penultimate quaver and the new text line begins on the last. In the first listening there is a PS peak on 2.5 and then on 2.875-3. By the final listening, there is a peak on 2.5 but this time there is a higher and more defined peak on 2.875. All ability groups contribute to the peaks.

The PE responses begin on 2.5, peak on 2.75 and fall towards 3. The EOP responses begin on 2.4375 and continue to a peak at 2.75. In general, therefore, responses coincide with the end of the first solo (and text) line and the beginning of the next. However, the spread of responses over the half bar during different listenings and by different listeners seems to relate to the distance between the resolution (or the expected location of the resolution) and the PS.

A similar response pattern is seen in the performances. Here the vocal line is more easily identified, so it could be expected that the responses for PE and PS would be more concentrated at the positions of 'actual' phrase ends and starts of the vocal line. The EOP responses begin to rise in the first beat of bar 2 and peak at 2.5, with a slightly higher response in for the Furtwängler than for the Leonhardt recording. The PE responses rise for 2.5 and peak at 2.75. However, the PS responses differ for the two performances. For the Furtwängler, the peak is at 2.875 while for the Leonhardt the peak is at 2.5. The tempo contours show that the peak of a small ritardando, and the end of a diminuendo coincides with 2.5. For the Furtwängler however, there is a slight acceleration through the bar and there is also a diminuendo to the half bar. However, the greatest difference between the two recordings is the underlying tempo; the Furtwängler recording is much slower. It seems, therefore, that listeners to the Furtwängler have time to hear and respond to both the PSs and PEs and, because of the tempo so far, are prepared for a temporal gap between them. In the Leonhardt recording however, they expect and hear the end of the phrase and immediately press the key again for the beginning of the next one even though they have not heard it yet.

Bar 4

Phrase end responses in bar 4 are spread but peak on 4.25. The highest PS peak is on 4.375. There may be two musical factors contributing to the relative inaccuracy. In the orchestral part, this is the first time in the movement that the cor anglais triplets have stopped and that they have reached the e^1. This could imply that a new beginning is starting. In the vocal part, bar 3 implies that f# will follow at the start of bar 4. This is 'closing' the opening high f# and so may be felt to be an end.

The largest EOP peak is on 3.625, coinciding with the second note of the 'closing motive' and the change of chord discussed above. There are two smaller peaks that coincide with chosen PEs on 4.25 and 5.

For the two performances, PE responses coincide with a peak on 4.25 and the PS peak on 4.375 with responses to the Leonhardt anticipating the PS but less so than in bar 2. The EOP responses are similar to those for the MIDI with peaks on 3.5 and 4.

Bar 5

In bar 5 there is a peak on the bar line (where there is also a change in bass line, and the end of cor anglais' held third), a higher one on 5.5 (where there is a re-entry of the cor anglais semi-quaver triplets) and another on 5.875 (where the vocal line begins again), with a spread of responses between the last two. As in the previous positions, the PE responses are higher and more limited in area than the PS. The PE responses are on 5.625-5.75 and the EOP responses peak on 5 and 5.375. The change of harmony onto the f#7 chord occurs on 5 and the c# of the vocal line has been waiting to be resolved for some time.

This is the first position in which there is a more even spread of responses for the Furtwängler. There is a high response both on 5.5 and at the end of the bar. This could be because of the accompaniment (the return of the triplet semi-quavers in the cor-anglais) or because of the strong harmonic resolution. The rest of the PS and PE responses follow much the same trend as previous positions. The EOP responses show a rise to the highest response so far for the Furtwängler responses on 5.5, occurring with the clearest harmonic resolution so far. The tempo contours show a lengthening at this position (section 10.2.4).

Bar 7

There is a peak of PS responses on bar 7 and higher ones on 7.5 and 7.625 but these are all lower than the peaks than the responses so far in the piece. This bar also has low responses for PE for which the clearest peak is on 7.375. By final listening, the peak of the PS responses is on 7.625, which coincides with the beginning of the text, and follows a rest. In this way, it is much like the phrase start in bar 4 but it is less popular. This may be partly because of the structure of the previous phrase, which is subdivided into subsections through internal rests and because the next section (bars 7 – 9) acts as a close of the first half of the movement (following the text structure discussed above) with its relatively clear harmonic progression and use of the closing motive. In this position there seems to be a difference between DL, M and N.[85] The DL and M peak on 7.5 and 7.625, while the N peak on 7.25, 7.875 and 8. The rest, among the other features in the middle of the bar 7 may not influence N, though another reason may be that N press the key again immediately after they lift it marking the PE.

In the performances, Leonhardt seems to follow this interpretation by continuing a diminuendo from bar 7 until bar 9, while in the Furtwängler there is a decrease and increase in intensity in bar 7. This is reflected in the listeners' responses with a higher proportion of listeners identifying PS and EOP in bar 7 in the Furtwängler than in the Leonhardt recording.

Bar 8

There is a small peak on bar 8 which strongest in first listening and reduces in the next two. Reasons for PS identification here include: one of the few non-chromatic chords of the movement is here and it is e minor, and there is a change in the cor anglais' melodic material. However, by the last hearing other factors seem to be overriding for the listeners; this is the middle of the vocal melody and by the end of the bar there is preparation for the next PE.

Bar 9

There is a small peak on 9.125 (with the highest peak in the DL) but a larger peak on 9.375 (with the highest peak in the M), the latter coinciding with the start of the vocal line. There is smaller response in the N who also respond on 9.5. It may be that they coincide again with the cor-anglais change rather than with the vocal 'line'. A similar pattern is seen for the performances as before. The same applies here as described for bar 5. This time, however, the vocal rest is longer and the tonal arrival is clearer.

Bar 10

There is a PS response on 10, 10.5 and 10.875. Most DL and M chose on 10.875 and the following beat, coinciding with the vocal line, while most N chose on 10.5, coinciding with the cor anglais. A similar trend is seen in response to the performances for similar reasons as previous positions.

---

[85] DL – degree-level, M – musicians, N – non-musicians, see chapter 3, section 3.3.3.3.

Bar 11

The peaks for PS are on 11.5 (mainly N) and 11.875 (mainly DL and M). Again, a similar trend is seen in response to the performances for similar reasons as previous positions. The smaller responses on 11.5 (and 10.5) responses may be related to changes in texture.

Bar 13

The largest proportion of listeners identify a PE during the last note of the movement. A small number of listeners also identify a PS on 13.125 and a PE in the beats preceding this. However, overall, it seems that, like the beginning and unlike the rest of the piece, when it comes to the end of the piece, it is the end of the piece rather than the end of the melody is chosen as the PE.

10.2.3.2 Discussion of listeners' responses

General Discussion

For the most part, the areas of greatest response coincide with the vocal lines (settings of the text lines) and the rests between them. However, the differences in responses, such as between the PE responses of bars 2 and 4 and the limited response at bar 7, indicate that rests in the vocal line are not the only feature that listeners are responding to.

The MIDI PE responses are clearer than the PS ones, though for both the MIDI and performance responses there is a spread of responses within areas around many positions (chapters 3 and 4). This may be partly because before a PE, there is a preparation (expectation-generating elements such as the melodic and harmonic characteristics section 10.2.2), so listeners are prepared for the PE when it arrives. In many other pieces, PSs follow PEs immediately, generating a general expectation for the PS to occur immediately here too. Listeners may therefore expect a PS but are not given a musical preparation for its actual arrival. It may also be that listeners, in this experimental structure, feel uncomfortable when they are not marking a phrase.

The ambiguities also encountered in other pieces explored in this study are added to here with the contradictions between the vocal and the orchestral parts. In most cases, the orchestral phrases start three quavers before the vocal ones. These three quavers can be heard as an introduction, or the vocal line can be heard as a continuation. In some cases, like in bar 1, the relationship is clear. In others, as in bars 2 and 4, there are few indications from the orchestral part of a true PS but there is enough ambiguity to cause difficulties. The listeners chose one position or the other. Some specified that they were trying to respond to one or the other part in different listenings. Graph 3.6.1.5, appendix 3.6.1 shows a comparison between the key presses for positions within each bar that may be associated with orchestral or vocal PSs. There is a larger proportion (to varying degrees) choosing the 'vocal' starts than the 'orchestral' starts. There are four exceptions: 1) in bar 1, a larger proportion chose the beginning of the piece, 2) in bar 5 the number of responses is the same on 5.5 (orchestral PS) and 5.875 (vocal PS). Both are preceded by a rest in their own part, both involve a change, and both are part of the same harmony, 3) in bar 9 again larger proportion choose the orchestral PS (9) than the vocal one (9.375). The resolution of the clearest cadence in the movement occurs on bar 9. The phrase boundary is an elided one, so the PS also coincides with this strong resolution. The vocal part simply continues from within this harmony, and 4) at the end there is no vocal part.

'Learning'

The responses indicate that there are some cases in which there is a relatively 'systematic' change from confusion to decision over the three listenings, such as bar 7, while there are others, for which clarity turns to confusion such as bar 4. The difference between these two bars may indicate the reasons for such different trends.

Bar 7 seems to be one of the more 'uncertain' phrase boundaries:
1) It follows a text line that has been, for the only time in the movement, fragmented so the 'pattern' that had been established is now broken
2) It is the first time that the PS is on this metrical position – not just a weak beat but an 'after beat' (7.625)
3) It is not near a harmonic resolution or an expectation for one, indeed for some it is not a phrase boundary at all (as supported by the relatively small level of response in this bar).

However, there are some features that do indicate a PE: the rhythm is the same as that of the PE of bar 2, this is followed by a rest and if the rests in bar 6 are seen as part of the phrase, this is as close as we get to phrase that is of equal length to the previous one. It seems that those who do respond here, though they do not expect a phrase boundary here in the first listenings, they 'learn' about the presence of some features and respond here in later listenings.

On the other hand, bar 4, with its high response on 4.375 in the first listening, and more spread responses in later ones, seems to have the opposite effect. Here, the PE seems to be expected in the first listening: 1) it is harmonically relatively well prepared, but not resolved 2) it has a melodic close 3) it has a change in the cor anglais part. However, having heard the rest of the piece, the ambiguities, and the different possibilities of types of phrase boundaries, it seems that listeners' responses are diminish in later listenings.

It seems that there is not an 'automatic' systematic pattern of changing responses over listenings resulting purely from more exposure to the same piece (chapter 8 and, for example, Deliège, 1998). Instead, there are different patterns of changes for different positions that seem to be driven by the musical features encountered.

'Musical experience'

The MIDI results indicate that all the groups tend, for the most part, to respond to similar positions. However, there are positions where there is a very small response for N while there are peaks for DL and M, such as PSs in bars 5 and 7 and there are some differences. Many of the differences in response are when there are changes in the cor-anglais material and these changes do not coincide with the solo line boundaries. From discussion with listeners, it seems that more N tried to concentrate only on the orchestral parts, while more DL and M tried to concentrate on the vocal line. This may be because the vocal line is not so much of a 'line' but in places disjointed and unexpected, while the cor-anglais move only in tones and semitones in a simpler manner with clear, recurring rhythmic and melodic patterns. In this MIDI rendition, there was relatively little difference between the solo line and the others. The responses of the three groups are much more similar for the performances in which the solo line is much clearer.

It seems, therefore, that musical experience did affect the listeners' use of features cues for the MIDI rendition. However, the difference almost disappears for the performance responses.

10.2.4 Studies of performances

10.2.4.1 Tempo Contours

The performances have different underlying tempi, different degrees and directions of change (graphs, 3.6.1.6, appendix 3.6 and chapter 7). Furtwängler's recording is the slowest and changes from the 'underlying tempo' are generally lengthening. Leonhardt's recording is faster and changes are generally to shorten the notes further. There are different types of tempo and intensity changes. To begin to distinguish between different types, two of each are discussed here.

Local accentuation by lengthening

Some single beats are lengthened, without lengthening of surrounding beats. This often coincides with the words set to relatively long notes (table 3.6.1.7a, appendix 3.6 and section 10.2.2.1). In general, few responses coincide with position with such lengthenings except the upbeat to 7 in response to both recordings and in 12 – 12.5 for the Leonhardt, first session (appendix 3.6.1). The upbeat to 7 coincides with a mid-line 'comma' in the text, and in the music, this upbeat follows a rest. The response for 12-12.5 in the Leonhardt is very small, is only in the first session, and is only by the musicians. Overall, this specific beat lengthening does not seem to affect phrase identification.

Gradual lengthening (ritardandi)

Other beats are lengthened as part of more gradual beat lengthening (table 3.6.1.7b, appendix 3.6). Some, especially the ends of the lengthening, coincide with the end of text lines and listeners' responses (in bold in the table) or more rarely with the rests between them (italic) and a small number coincide with the individual words described above (underlined). In some cases, listeners' responses coincide with these ritardandi. However, this is not consistent for all the phrases identified by listeners or indicated by the text. When there are ritardandi, they are to different degrees and in different positions. The positions of ritardandi coincide with responses. However, there are also positions that do not have such ritardando that also have high responses.

It seems that even though this kind of ritardando does occur in music of this era (contrary to the description in Friberg and Battel, 2002, p. 20, see also chapters 7 and 10.3), it does not occur at every position and is not necessary for listeners to identify PS, PE or EOP. It does not even seem to affect the proportions and does not occur in the run up to the big phrase boundary of bars 8-9.

10.2.4.2 Dynamics/Intensity Contours

Accents

Table 3.6.1.6a, appendix 3.6.1 shows the individual words that were lengthened in performance and whether there was an increase in intensity with them. Comparison of tempo and dynamic contours reveals that some words are both sung more loudly and lengthened, whilst for others, there is only one performance feature. Many of these words are set to notes that are amongst the highest in the piece and this may contribute to their higher intensity. However, the c# for **Kost**barkeit is sung more loudly, while the c for Tränen is not, indicating that pitch height is not the only intensity determining factor.

Gradual decrease in intensity (diminuendo)

The areas over which there are gradual decreases in intensity over several notes (diminuendo) are shown in table 3.6.1.7b, appendix 3.6.1. The span is from the peak of the intensity to the following trough. The table shows that there are several areas in which there is a diminuendo for all performances some of which coincide with listeners' responses. Broadly, the gradual diminuendi often (but not always) follow the rule of 'the higher the louder', or rather, the lower the quieter. For the Leonhardt and Cleobury performances there seems to be high coincidence between areas of decrease in intensity and in tempo (graphs 6.1.6-9, and tables 3.6.1.6b and 7b, appendix 3.6.1). For Furtwängler, however, the two seem mutually exclusive: either there is a diminuendo or a ritardando.

10.2.5 Summary of the musical analysis, results of listening and performance studies and identification of musical features

The analysis and discussion suggests that there are a number of musical and performance features that may be contributing to PS, EOP, and PE identification in the listening studies.

10.2.5.1 The text and listener responses

In most cases, PSs/PEs chosen by listeners are in the same areas as the text-lines starts and ends. However, when there is some punctuation within the text line (such as in line 4),[86] further possible candidates for phrase division arise.

As with the Schubert Lieder (chapter 2), the musical structure tends to closely follow the text structure, so the phrase structure can, to a certain extent, be predicted from that of the text. However, where there are areas of ambiguity in the text, there can be ambiguity in the music (that may be reduced in retrospect), and moreover, the musical structure can sometimes 'contradict' at least the superficial text structure. In addition, though the vocal and text lines coincide, the accompaniment in this movement does not always coincide with the vocal lines (and therefore the text lines). Relying on text would mean that crucial phrases would be missed (such as the important bars 8-9).

The rhythmic structure of the text seems to have been followed closely, particularly in the location of the up-beat starts, and strong/weak-beat PEs and the musical structure seems to coincide with the rhyming scheme and meaning.

All this indicates that though the text can be used in phrase structure identification, it may not be completely reliable if treated 'automatically'.

10.2.5.2 The 'Harmony'

The complex harmonic structure of this movement, and the mismatch between voice and orchestra, is reflected in the MIDI and performance responses. The responses show that expectation of resolution seems to play a large part in phrase identification; a complete resolution is not necessary for PS, PE or EOP identification. Another view could be that harmony is irrelevant and that listeners are responding only to other cues. However, this is not supported for the following reasons: the responses here are clear but lower than for other pieces, indicating that the phrases are not as clear as in the other movements. However, the other cues, such as the rests between the phrases, are very clear and attract listeners' responses. Moreover, where the harmonic close is clearest (bar 9), the response rises and coincides with the harmonic resolution as well as the vocal end

[86] or even some that is implied by the setting (such as Fleisch und).

and start. It seems therefore, that the PE/PS responses may be at least in part related to harmonic structure and the expectations generated by it (EOP).

10.2.5.3 Closing Melodic Motive

The melodic closing motive seems to develop from bars 3-4, becomes explicit by bars 7-8 and repeated in bars 10 and 12-13. It may help signal the approaching PE and therefore contribute to EOP and PE responses.

10.2.5.4 Metrical Structure

The phrases identified in the piece hover around the bar and half bar (with their constantly shifting importance). All but one of the PSs is on a quaver-long upbeat, while bar 7 starts on a three-note upbeat. This is an example of the clear distinction between the phrase and metrical structures but the close relationship between them; the phrase structure clarifies which part of the bar is more important, while the metrical structure shows which notes near the beginning or the end of the phrase are important and therefore the character of the phrase (including, up-beat, weak close, strong close). Having established this relationship between metrical structure and phrases, listeners may begin to expect PEs and PSs in particular metrical positions.

10.2.5.5 Texture

The responses indicate that the changes in texture (resulting from the cor-anglais motives) sometimes coincide with a relatively low response even when the other features do not all coincide exactly at that position. This could be because some listeners, especially to the MIDI rendition, took the cor-anglais to be the melody line (as listeners reported). It could also be because sometimes the change in texture is a strong cue, regardless of what else is occurring. It seems that changing texture can coincide with responses but that when not confirmed by other features, these positions are relatively weak candidates for phrase boundaries.

10.2.5.6 Rests

Rests (or breaths) separate vocal (and text) phrases. As will become clear responses coincide with rests more systematically in this piece than the other case-study pieces. Though relying on rests to identify phrase boundaries would be mostly successful in predicting the listeners' responses, this study shows that, even here, other musical features contribute to the different types of responses.

10.2.5.7 Features, their positions and combinations

As with the other case-study pieces, the musical features of the piece were identified on the basis of the above music analysis (section 10.2.2, and graph

3.6.1.10, appendix 3.6.1). Here they are described according to bar locations beginning with the most feature intensive areas.

Bar 4

The first feature intensive area is on bar 4 and follows preparatory or predictive features (cadential progression in harmony and voice, change in underlying rhythm, voice-leading), instantaneous ones (pitch jump in the accompaniment, bar line, a retrospective one (change in texture), and ritardandi in the recordings. From here there is an expectation for a PS. Those listening mainly to the accompaniment may have identified a PS on bar 4, or may wait until bar 5.5 (see below). Those listening to the voice, wait until after the rest in the vocal line (bar 4.375) and some respond one quaver later, on the metrically stronger half-bar.

Bar 10

The next feature intensive area is on bar 10. Again it follows preparatory features (following cadential progression and voice-leading in the voice), instantaneous ones (following pitch jumps and long notes/rests), a retrospective one (change in texture) and ritardandi in both recordings. Again, from here onwards there is an expectation for a new PS. Those listening to the accompaniment identify it on 10.5. Those listening to the voice identify 10.875 or in the following (down)beat (the stronger metrical position).

The beginning of bar 11 also has five features. There are no preparatory features but instantaneous ones (pitch jumps in both the voice and the accompaniment), retrospective ones (change in texture), bar line and ritardandi in both performances. There is also a small PS response for the accompaniment on 11.5.

Bar 5

The next feature intensive area is on bar 5.5. Here there are less preparatory features (cadential preparation), two instantaneous ones (pitch jump and long notes) the same retrospective one (change in texture) and lengthening in the Furtwängler. Here the distribution of responses is different from that of bar 4. In bar 4 the vocal phrase is relatively clear while the accompaniment arrives at a long note. Half way through bar 5 the triplets of the accompaniment return and there is a clearer PS for those listening to the accompaniment. Those listening to the voice, wait for 5.875. Just like in the other positions, some respond on the metrically stronger first beat of bar 6.

Bar 9

The next feature intensive area, also with five features, is on bar 9 with more preparatory features (following increase in underlying rhythm, following cadential progression) an instantaneous feature (pitch jump) and bar line and four bar template. This area has the longest rest for the vocal line. Some listeners, having heard the end of the vocal phrase in bar 8.5, mark the PE there and then press the key immediately, regardless of whether or not the next phrase has begun. Those listening to the accompaniment, follow the features leading to bar 9 and respond on bar 9. Those listening primarily to the vocal line wait for the re-entry of the voice on 9.375. Again, there are those who wait for the metrically strong half-bar before they press the key.

There are 4 features on bar 8 (following voice-leading and cadential progression, change in texture, bar line and a change in the accompaniment). There is a small response here but the PE/PSs on bar 7.5 and around bar 9 are more popular.

Bar 13

The last bar starts with 4 features (the end of the vocal line) and 4 further features half way through. The start of the bar has two preparatory features (following voice-leading and cadential progression) and bar line and four-bar template, while the middle of the bar has two preparatory features (following a cadential progression in the accompaniment, following change in underlying rhythm) and the instantaneous pitch jump and change in texture. There are some who respond on 13.25: this is the beat after the vocal line finishes (so they could be expecting the next phrase) and is the beginning of the accompaniment's codetta.

And Bar 2

This covers almost all the areas that have high responses (section 10.2.3.1). Only bar 2 and the beginning have been excluded (for the latter see 10.2.3.1). The area of bar 2 has very few features, so why is there such as high response by the listeners? The features present are: Following cadential preparation, supported by voice-leading on bar 2.25 both in the voice and in the cor anglais, especially the second, on 2.5 and a rest on 2.75 with the voice starting on 2.875. It seems that some key features, do not need others to elicit a response.

10.2.6 Major characteristics of phrasing features in the Bach Passion

There is very little repetition in the vocal line so unlike in the Wagner, Brahms and to some extent Bach Suite, there is little effect of repetition. This, combined with the temporal gap between the PE and PS (mainly in the vocal line but also in the accompaniment), results in a lack of confirmation features. Instead, it seems that there is reaction to preparation for PEs and changes in texture or rests. The piece is harmonically complex, so long term connections are difficult to make, but preparation features seem to include local harmonic preparations. It seems, therefore, that local harmonic and melodic preparation features (in the voice and accompaniment), supported by rests and changes in register, are the main features.

The accompaniment stresses the metrical structure (the bar and the half bar) by beginning and ending its triplet sections, long notes, and cadences on these positions. However, none of the vocal phrases begin on a strong beat (bar or half bar). As a result, there is a strong pulse (and "meta-pulse") underlying the piece but the vocal phrases either anticipate or are delayed in relation to it. The importance of the pulse is seen in the responses of some listeners who mark the PSs (and PEs) in line with the nearest main beat rather than the off-beat.

The structures of the accompaniment and voice seem to compete with yet support each other. The phrases of these two overlap most of the time, leaving listeners with a dilemma: to treat the two as a whole, or to listen to one part. In the case of the former, there are few clear phrase boundaries where the two coincide or support each other, such as that of bar 9. In the case of the latter, those who listen to the accompaniment find it relatively easy to dissociate from the vocal line, as the PSs and PEs coincide with metrical accents. Those listening to the vocal line could rely simply on the frequent rests in the piece. This would result in all such PSs having a stable number of responses. However, this is not the case. Instead, in places where there are additional features (both from the vocal line and the accompaniment) there is a higher response.

The positions of response to the different performances are similar; at any position that has a response to one recording there is a response to the other. However, the proportions of responses sometimes differ, reflecting more subtle differences between the performances.

For each area there is usually a pair of peaks of responses with a smaller number of further responses around them. For each pair there may be a different weighting of response. For example, in bars 2.75, 4.25, 7.5, 10.75 and 11.75 quavers 1 and 2, there is a higher response on the earlier quaver for the Leonhardt recording and on the later quaver for the Furtwängler. The Leonhardt responses coincide with the accompaniment and those for the Furtwängler coincide with the vocal line. The proportion of responses on bars 10.5 and 11.5 are almost the same.

The proportion of responses to both positions in bars 5 and 9 are equal. Here there are two clear options: the triplets of the accompaniment restart after their first rest in bar 5 and then the voice re-enters at the end of bar 5. In bar 9 the PS may be on the bar line or on 9.375. In these positions, where the harmonic, melodic and thematic information is so clear, the performance features do not seem to have different effect on responses.

The gaps between PEs and PS, and the mismatch between vocal and orchestral phrase parts, lead to the spread of responses. It seems that there is no one 'correct' position for a PE or PS, let alone boundary, but rather an area in which there are several possibilities, and more than one may be identified by listeners. This explains in more musical detail the observation of phrase-part areas (chapters 3, 4, 11 and 13).

10.3 Static harmony, clear general divisions and local ambiguities: Prelude of the Suite in C minor BWV 1011 by J.S. Bach (Bach Suite)

10.3.1 Introduction

Bach's music for unaccompanied violin and 'cello is famous for the ways that a single instrument produces different voices to form a polyphonic texture and that harmonic information is both given in polyphonic sections and implied in monophonic ones.

Bars 1-7 of the prelude of the Suite in C minor for 'cello BWV 1011 (Bach Suite) provide an opening statement characteristic of a prelude, establishing and emphasising the home key for several bars. They establish the motivic and rhythmic information that will form the basis for development and a metrical structure, though this is already challenged in bar 3. These characteristics are of particular importance for the phrase structure and are discussed here. The graphs in appendix 3.6.2 show the listeners' and annotators' responses and tempo and intensity contours.

10.3.2. The piece

10.3.2.1 Harmony

The tonic pedal that dominates this excerpt establishes the tonic. It is re-sounded every bar or bar and a half, each time playing an important structural rôle. Being the lowest open-string of the 'cello, it sounds throughout the section. Over the pedal there is a relatively weak harmonic move of broadly I (bars 1 – 3) to IV (bars 4 – 7). As the C pedal continues throughout and has the bass position, there is little chance of the c – f being felt as a strong V – I (in f) harmonic progression. Within this there is a chordal progression of:

Bars	Chords per bar
1	I
2	b dim (acts as V)
3	I
4	g dim
5	iv^{c}
6	i^{4-3}
7	Iv

The opening three bars establish the home key; the opening octave C is followed by the melodic minor scale from the dominant upward which highlights the minor third. The diminished chord on b followed by the return to i in bar 3 are all features that contribute to this establishment. Bars 4 – 5 introduce iv and bars 6 – 7 reiterate it but there is no escaping the underlying c minor harmony. Overall, however, the underlying harmony is static; for Rothstein (1989) there could be not be a complete phrase in the section (see also chapter 9).

10.3.2.2 Motives

The musical full stop (and the pedal)

In three of the five positions at which the low C is repeated, the C is reached through a leap down of an octave, a minor sixth, and an octave and a fourth on the second beats of bars 3, 5 and 7 respectively. In each, the rhythm is the same: two crotchets that are longer than the preceding notes and accompany the melody line. For each, the low C is an extension of the crotchet that precedes it. This is clear the first time it is heard (bar 3); the melody e*b* is held for a minim while the low C is played. Minims are not written in the later repetitions probably because it is not possible to play the other melody notes and the low C simultaneously on the cello, but the effect is the same. The rhythmic change, the return to the pedal note, and the position in the bar (an extension of the first beat) all give these positions a closing character.

For the PE MIDI responses, the most popular positions are bars 3.25, 5.25, and 7.25 (graph 3.6.2.1, appendix 3.6). Being the last note heard, the popularity of the last is not surprising. The only other PE that was chosen by half or more listeners and annotators (at least in the final listening) was bar 2. This is the end of the opening semi-quaver run, the first long note since the opening, and the first time the C is repeated. All these factors may contribute to its identification.

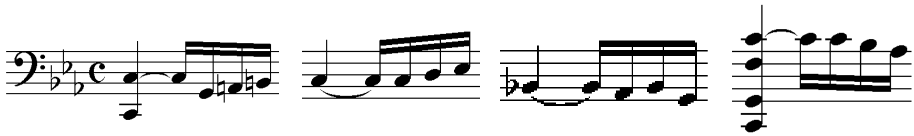
The only other repetition of the C is on bar 6. While 24% of listeners indicate a PE on the first note of this bar, there are reasons why this is not a clear PE. The C is part of a C^{4-3} progression that is not resolved until 6.5. It is also part of the opening motive so, if anything, it denotes an opening.

The above view sees the low C as connected to the previous note rather than starting something new. However, the listeners' responses indicate that this connection may not be so clear or, at least, predictable. For example, a clear resolution is reached on bar 3 and, slightly less strongly on bars 5 and 7. On first listening, 47% of listeners choose a PE on 3 and 3.25. By the final listening PE indications on bar 3 drop to 21% and those on 3.25 rise to 61%. Some listeners lift their finger and press again on the same beat. This could be because they identify the PE and PS on the same beat or because they are anticipating the PS. Similar patterns are seen in bars 5 and 7.

Each time the harmonic arrival and the full stop are spread over two beats: the arrival and its prolongation.

A rhythmic motive and its association with a beginning

The opening bar not only establishes the harmonic context but also presents one of the main motives of the piece (bar 1). Within the opening section the first two beats of the rhythmic motive are repeated three times.



Bar 1 Bar 3.5 Bar 4.5 Bar 6

The first time the motive is repeated (bar 3.5) it follows the crotchet leap back down to the pedal c (the musical full stop) and returns at the same pitch as the opening. Indeed it would be possible to start the piece again here. The connection with the opening (beginning), and the preceding return to the I (close) are both features that contribute to the indication that a new beginning starts in bar 3.5. A majority of listeners identified this position as a PS (and PE) (graphs 3.6.2.2-4, appendix 3.6.2).

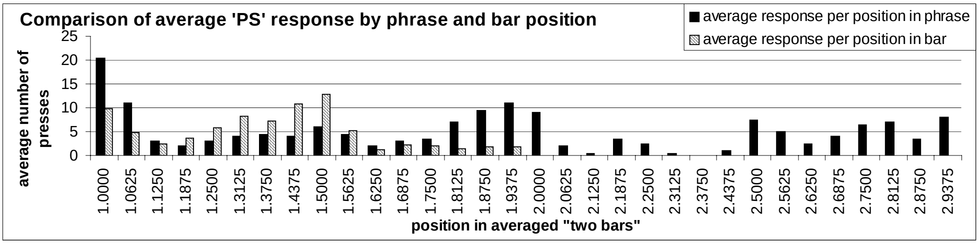
Having heard the motive twice in positions of a 'beginning', the motive now has this association. The next time the motive is heard (on 4.5) it is in its most varied form, yet even in the final listening some listeners respond here.

The last time the motive is heard (bar 6), it returns to c and the first beat of the bar. Again, it does not directly follow the falling crotchet motif. However, listeners respond here too. This motivic analysis is loosely based on the Rétian approach (chapter 9.1.1).

10.3.2.3 Metrical Structure

The positions of the full stops and the rhythmic motive together create a strong sense of an end and start in terms of the phrase and in terms of the metrical structure. As these move between leading to and from the first and third beats, the first and third beats of the bar become equated. As a result, there is a sense in which the opening could return on bar 3.5 despite the fact that this is half way through the bar (section 10.3.2.2). Despite the strength of bar 3.5, there are still some listeners who, having chosen bar 1.0 and 2.0 as PS, also choose bar 4 over 3.5. However, relatively few listeners do so in the last hearing (graphs 3.6.2.5, appendix 3.6.2). It could be concluded from the tendency to identify 4 over 3.5 that at least on first listening, the PS is simply the beginning of the bar. To compare these beginnings more closely, the total number of MIDI phrase responses in each position that was not on a bar-lines (*phrases*, bars 3.5, 5.5) and those that did occur with bar lines marked in the score (*bars*) were averaged and compared. The results are presented in graph 10.3.2.3 and show that responses at *phrases* were more than double those at *bars* suggesting that bar lines are not clear indicators of phrase starts for listeners.

Graph 10.3.2.3



10.3.2.4 Texture

Although this piece is monophonic, the number of concurrent voices changes throughout: a single line, chords and "two lines". The listeners' responses coincide for the most part with last two: either a chord (usually PS, such as bar 1), or two parts (more often PE, such as bars 3 and 7). Like with the motivic repetition however, this feature occurs also in positions that otherwise do not have such clear indications of PSs or PEs (such as bars 4.5 and 6). This feature strengthens positions with other musical features (such as bars 1, 3 and 7) and, like with the motivic repetitions, may also be one of the reasons that a small proportion of listeners indicate PSs at positions that otherwise do not have such clear indications (such as bars 4.5 and 6).

10.3.3. Listeners' and Written Responses

Here the areas of high response (chapters 3 and 4) are discussed. The graphs in appendix 3.6.2 show the listeners' and annotators' responses and tempo and intensity contours.

Bar 1

The prelude is for one instrument (that accompanies itself), so, unlike for the Bach Passion, there is no 'introduction'. The vast majority listeners PS responses are with the opening octave for all three renditions (MIDI, Gendron and Rostropovich) and in the written responses. However, a small group of both of listeners and annotators respond on the second note and have equivalent responses for later bars. If only the listeners' responses were analysed it may be concluded that these responses were 'late' i.e. mistakes. However, as annotators gave responses in the same place, it seems that these are not mistakes at all. Instead, they seem to be systematic choices (at least in the cases of the annotators). This response may be because the bass part, and anything simultaneous with it, is treated separately and therefore not included as a PS.

Bar 2

On bar 2, the EOP, PE and PS peaks coincide (with a small response of EOP in the preceding bar). The run of semiquavers, followed by the final four of the bar that hover around a close, followed by the change of texture to a full chord (returning to the pedal C) on a long note, all seem to contribute to a PE/PS response here. Although this is an important chord (the first bar may almost be an upbeat to it), this seems to have a different character from later PSs in the piece, which have a clear, repeating pattern. This can be seen in the relative size of the MIDI responses; the response in bar 2 is only higher than that of bar 4, which is even more problematic. The differences between the performances here illustrate some of the different interpretations at this position (graphs 3.6.2.7-10, appendix 3.6.2)

Gendron's performance plays down this position. The changes in tempo and dynamics around bar 2 are much less perceptible than those in Rostropovich's recording, in which the last beat of bar 1 and the first two beats of bar 2 are lengthened. The differences in impact of these different performances can be seen in the higher responses for PS and PE and earlier EOP responses for the Rostropovich recording than for the Gendron. This is an example in which the music can be interpreted in more than one way and in which the performances can affect the phrase interpretation.

In the written responses 2 and 2.375 are chosen. This is one of the few positions for which there is a difference in proportion between the MIDI responses and those of the written responses. For the MIDI there is an almost equal level of response at these two positions. In the written version, it is easier to show accurately where the phrase start is intended and though both positions are chosen a greater proportion choose 2.375.

Bar 3

The next high MIDI response area is in bar 3 and is one of the most popular positions in the written and performance responses, the position of the 'musical full stop', preceded by a descending line back to the tonic and followed by a repetition of the opening motive. There are some PE MIDI responses on 3 but the majority are on 3.25. The PS responses begin on 3 but peak on 3.5. In a situation similar to that in the Bach Passion, the PS keys are pressed from the first possible moment that the previous phrase could have ended which falls in the middle of the PE 'extension'. The EOP starts at 2.75 (the $\hat{3}$ of the $\hat{6}$-$\hat{5}$-$\hat{4}$-$\hat{3}$-$\hat{2}$-$\hat{1}$ descent) and peaks again on 3.

The PS responses to both performances peak on 3.5. In the responses to performances the PS peaks are on 3.5 for both. However, the Rostropovich responses also peak on 3 (first session only) and 3.25 (both sessions). For the PE the Gendron responses peak on 3.5 but also have responses on 3 and the Rostropovich responses have a peak on 3 and 3.25 in the first session and only on 3.25 in the second. The two recordings are different both in their underlying tempo (the Gendron is faster than the Rostropovich, see below section 10.3.4.1) and the articulation of the notes – Gendron leads to the half bar, while Rostropovich accentuates each note.

Here and for other positions, the responses for the PSs are sharper than PEs and the EOPs, particularly in the performances (graphs 3.6.2.2 and 3.6.2.4, appendix 3.6.2). This difference between PS and PE is not seen in all pieces (Mozart Sonata, chapter 10.5) and coincides with the differences in the durations of the 'impact' (chapters 11 and 12) of musical features. Here, the arrival on the harmonic resolution and its prolongation (accentuated by performance features) in the musical full stop last two beats, while the PS begins 'instantaneously' on the next beat.

The PEs of bars 3 and 5 are spread over 3 beats in the listeners' and written responses. Each of the beats may be chosen for different reasons. As in the case of the few 'late' PS responses (see discussion concerning Bar 1, above), the bass may be treated separately. Here the separation is clearer as the C onset is alone at the end of each phrase. In addition, the resolution is reached on the first beat of the bar, and it may be for these reasons that some identify PEs on bars 3, 5 and 7. For those that include the bass note as part of the texture to be considered, the resolution is finally reached on 3.25, 5.25 and 7.25. For others, the end does not arrive until bars 3.5 and 5.5.

Bar 4

There is a relatively small MIDI response on bar 4. Like in bar 2, PS, PE and EOP responses coincide, though the EOP responses also have a small peak on the preceding beat. The response to the performances is delayed to the next note (4.375). In the written responses, positions 4.375 and 4.5 are chosen by a small number of musicians (1 and 4 respectively). Bar 4 contains several features that coincide with other PSs in this piece; the downbeat position in the bar, the dotted-crotchet quaver rhythm of bar 2 and also has a sense of arrival following the c d e*b* semiquavers. However, this position does not have a repetition of the opening motive and does not immediately follow the musical full stop. This feature combination therefore has contradicting effects and the proportion of responses here is relatively low.

Bar 5

Bar 5 is very similar to bar 3 both in terms of musical features and in terms of the high proportion and pattern of listeners' and written responses. Again, there is the 'musical full stop' and metrically the two are in parallel positions. Again, the full stop is followed by a motive that could have signified a PS earlier in the piece. This time, however, the theme is from the 'weaker' position of bar 2 and is immediately followed by the 'stronger' opening motive (clearly on C) coinciding with some responses on bar 6. As in the upbeats to bars 2 and 3 Gendron lengthens the upbeat to 5, which seems to affect the EOP responses (see discussion of **Bar 3**, above).

Bars 6-7

There is an EOP response on 6.75 and a PE response on the following two beats in the listeners' and written responses.

10.3.4 Studies of Performances

10.3.4.1 Tempo Contours

In comparison with the listeners' responses to MIDI many of the positions lengthened here are chosen by listeners as PE (such as 3.25), or precede positions chosen (such as 1.875). In general, the responses to the performances are similar. There are, however, some details which differ. In some cases, these coincide with the lengthenings identified in the performance contour. For example, Gendron lengthens 2.75-2.875 while Rostropovich does not. The PE and PS responses on 3 are higher for Gendron than for Rostropovich. Conversely, Rostropovich lengthens 3.25 and 4.375 while Gendron does not. The responses on these two positions are higher in both listenings for Rostropovich than Gendron. Rostropovich also lengthens 5.25 and 5.875-6 but less dramatically and this time the difference in responses is not large between listenings. The differences in responses to the two recordings as a whole are not significant (chapters 3 and 4). However, the individual differences indicate that the changes in beat length do affect the proportion of PS and PE responses.

10.3.4.2 Intensity Contours

Overall the Gendron intensity contour seems to descend steadily across the whole section and more dramatically so at the very end, while the Rostropovich intensity contour as a high start (with the first chord) and then for the most part remains within the same range with some peaks and troughs within it. Gendron's minimum at bar 2 is longer and lower than that of Rostropovich, while that in bar

3 is more sudden and deeper for Gendron and more gradual for Rostropovich. This is reflected in the responses in that there are more earlier responses in the first listening to Rostropovich than to Gendron. Neither performer has such a clear minimum in bar 5. Gendron has a diminuendo at the end of the extract, while Rostropovich does not. This is reflected in the responses; for the Gendron there are more listeners expecting a new phrase start and the PE responses begin earlier than for Rostropovich.

It seems from the intensity contour and tempo graphs that the responses can related both to tempo and intensity change.

10.3.5 The features in view of listeners' responses,

Some analysts would say that bars 1-7 of this prelude form the beginning of a phrase that continues beyond the end of the excerpt because there has been no harmonic motion away from the tonic (Rothstein 1989). Despite this, when asked to make phrase decisions, listeners do so and the positions chosen coincide with distinct musical characteristics:

1. The 'full stop' of the low C
2. The preparation for the close before it
3. The beginning of the rhythmic motive

All consisting of harmonic and rhythmic, voice-leading, motivic and intervallic features.

In the features graph (graph 3.6.2.12, appendix 3.6.2), some the features have been marked as occurring on one position on the graph, while their effect, and the listeners' responses occur over an area (chapters 11 and 12). There are three areas of greatest response to the Bach Suite: bars (1) 3, 5 and (7). Two key characteristics are highlighted: the effect of elongating the end of a phrase (resulting in a metrical shift) and the effect of one harmony dominating the entire segment. Some of these are briefly highlighted.

Bar 3

By the beginning of bar 3 several preparatory features of cadential progression, melodic voice-leading and change in underlying rhythm have occurred and a long note is reached. Both performers lengthen bar 3.25. An inexact repeat of the start of the piece (and the motive) begins 3.5. Here, with the repeat of the beginning and following the clear conclusion of the previous phrase, the new phrase begins.

Bar 5

Similarly in bar 5, the phrase ends on beats 1-2. Bar 5.25 is lengthened by both performers and the phrase begins with a chord and following a long note half way through the bar. Although the features named at the start of the bar (bar line, four bar template, and chords) are different, the effect in terms of response is the same as in bar 3.

The metrical structure is played with; it is as though the bar line is shifted by half a bar from bar 3. Having shifted the accent structure, the full stop returns at equal length intervals making the phrase structure regular and therefore predictable.

The phrase start follows another prolonged phrase end. Some listeners, having lifted the key for the phrase end on bars 3, 5 and 7 or during their first two beats, immediately press the key again, not waiting for the "actual" phrase start.

Bar 6

A further trick is played on the start of bar 6 with the opening motive returning (supported by a clear chord). Despite these features, listeners do not identify a PS here. This is important as it shows that simple repetition of a rhythmic motive (an inexact repeat), even such a strong one, does not necessarily encourage PS identification. The features of previous PE and new beginning and the regularity of the phrase structure are strong enough to outweigh the features of bar 6.

Bars 2 and 4

There are also responses in bars 2 and 4 though in smaller proportions. In bar 2 this coincides with the change of motive and texture on the bar line and then half way through the bar. In bar 4, this coincides with the repetition of the rhythmic motive of the opening. The number of responses here, however, is relatively small.

10.3.6 Major characteristics of phrasing features in the Bach Suite

This discussions shows that the features identified through musical analysis coincide with listeners' identification of PS, PE and EOP. In the absence of clear harmonic motion away and back to the tonic, there are systematic responses which coincide with musical features that include the harmonic one of the reiteration of the tonic, repetition of both closing (full stop, which includes the tonal end and the rhythmic close) and repetition of opening motive.

This discussion also indicates that the features and the phrase parts are not limited in their duration or 'impact' to one position (chapters 11-13). Many of the features indicating the EOP and PE last more than one note and have an impact lasting several notes (including voice-leading, those included in the musical full stop) and those indicating PSs (including repetition) last more than one note but can have a more retrospective or immediate impact (especially once the piece is more familiar).

Furthermore, this discussion suggests that there are positions at which several features occur together or in succession and strengthen one another, and these are identified in the majority of responses. There are also positions with fewer features which are identified by fewer listeners. The smaller responses at these positions indicate that, whilst these are not clear phrase parts to the majority of listeners in all listenings, they do still suggest phrase parts.

In performance, the importance of these features may be reduced by suppressing these positions and accentuating others. This results in even fewer listener responses at these positions and more responses at the 'main' ones. It is in this way, i.e. in the degree rather than location of responses, that the listeners' responses to MIDI, performances and scores differ.

10.4 Clarity, and complexity, classic harmonic structures with melodic counterpoint: The Slow Movement of Mozart's Piano Sonata, K. 310 (Mozart Sonata)

10.4.1 Introduction

Mozart's music, and often his piano works, are referred to as prime examples of pieces from the eighteenth century in which the phrasing is particularly clear and foursquare (Keller 1965; Grave 1980).

Davis explains that one of the most important elements in music of the eighteenth century, and perhaps the most fundamental determinant of classical sonata-form, is tonality.[87] Harmonic rhythm[88] contributes to relatively small formal subdivisions and often provides significant articulations between themes and groups of themes.[89] It complements tonal rhythm; where tonality is stable (as in a primary theme) the harmonic rhythm shows maximum activity and differentiation; where tonality is in flux, the harmonic rhythm maintains stability and regularity in the smaller dimension (Davis 1966, pp. 25 and 35). In the classical language, 'interactions of rhythm and harmony typically involve the juxtaposition of tonic harmony (denoting stability) and the dominant (suggesting mobility)' (Grave 1980, p. 88).

---

[87] For example, 'tonal change to the dominant usually signals the entrance of the secondary thematic group in expositions; the return to the tonic establishes the beginning of recapitulations, and rapid modulation – faster tonal rhythm – characterises the excursions of the developments' (Davis, 1966, p. 25).
[88] the rate of root change in harmonic progressions (for example, Davis, 1966, p. 25).
[89] The different sections of a sonata form each have different characteristics including differences in harmonic-rhythmic structuring as well as melodic differences and changes of tessitura. The primary theme section is, for example, characterised by increasing speed: one or more acceleration designs, i.e. groupings of progressively smaller note values, often dominate (Davis, 1966, p. 27).

Grave compares Mozart's music to that of his contemporaries. In the latter, 'conventional features on which the style depends tend to be all to readily apparent: predictable, foursquare rhythms of phrasing; mechanical symmetries of sub-phrase, phrase and larger particles of structure; and ubiquitous cadences, in which habitual coordinations of harmony, rhythm and line yield a ready source of punctuation and closure. These standard materials aid the composer in fashioning large-scale, hierarchic designs marked by clarity of statement and architectural balance. But these same resources may also hinder the projection of a sweeping, goal-directed movement by which otherwise mechanical phrase separation might be overcome. Although Mozart draws on the available conventions of his day no less eagerly than do his contemporaries, those forces that might threaten continuity are somehow conquered, and long spans of spontaneous, goal directed energy are accomplished' (1980, p. 87).

The opening of the slow movement of Mozart's Sonata K. 310 (bars 1-8, Mozart Sonata) contains classic cadential progressions underpinning its structure and fulfils the role of an opening. Here the key, metre, and thematic material are established and can be developed and contrasted with other material later in the movement.

10.4.2 The piece

Harmony, Melody, Metrical Structure

F minor is established almost immediately: the upbeat, an f minor arpeggio, is followed by I – V progression in bar 1 which acts as the preparation for a perfect cadence. Instead of reaching Ia immediately, it reaches VI (an interrupted cadence, chapters 1, 9 and 12) with a 6-5, 4-3, suspension causing the resolution to fall on the weak (second) beat of bar 2, without reiteration of the bass note.

The anacrusis to bar 3 is extended by two demi–semiquavers (if compared to the opening) and leads into a bar dominated by V/V leading to a V in bar 4 with another 6-5, 4-3 suspension, delaying the resolution of the chord to the second (weak) beat of bar 4 (without reiteration of the bass). The tonic note is reached on the (weak) third beat and is sounded only briefly, without support in the bass and the melody immediately runs up a scale. Here different features of the melody can, in theory, suggest different locations for a boundary:

If the e and f of the 4.333 and 4.666 of the melody were of equal lengths, the phrase would end on the f. However, this voice-leading is not supported by the bass (which does not provide the tonic chord until bar 5) or by the rhythm of the melody (which gives the f only a demi-semiquaver). The rhythm and the explicit harmony lead to pause on the second beat (on an imperfect cadence).

The piece (and first phrase) started with an upbeat at the start of the last beat of the bar. The next phrase is expected to begin in the same position in the bar (i.e.

the f). This leaves two possibilities: either the previous phrase ends on bar 4.333 and the next starts on 4.666, or the previous phrase ends on the same beat as the next phrase begins (4.666). This illustrates the relative importance for the phrasing of the accompaniment (and explicit harmony), and rhythm and voice-leading of the melody.

The scale (4.666) forms the upbeat to an ornamented repeat of the opening, which is supported by the tonic again on bar 5. The left hand part here is identical to that of the first bar and a half, leading to the same interrupted cadence as bar 2. The arpeggio of the opening upbeat is elaborated and the scale steps are 'filled in' to form an almost complete scale. The melody of the bars 6-6.333 is modified; instead of the dotted quaver – demi-semiquaver – dotted quaver of bar 2 there are two quavers followed by a semiquaver. There is therefore, less of a pause on bar 6.333 than on bar 2.333.

In bar 7, the bass is also modified and the following bar and a half is a preparation for, and resolution of, a perfect cadence – the first complete perfect cadence of the movement (with the tonic in the melody and the bass). The tonic arrives on the strong (first) beat of bar 8. This is early in comparison to the positions of the earlier cadences. Here, also the goal of the voice-leading of the opening section is finally reached. Once the tonic is reached, the texture, theme and mood of the music changes. This is the end of the opening section of the movement and the end of the extract used in the current study.

Overall, there are four cadences (bars 2, 4, 6, 8), three of which are followed by a hiatus. That of bar 8 is most complete, that of bar 4 is next (with its own ambiguity discussed above). Bars 2 and 6 are interrupted cadences and continuation is expected, with 6 being even less complete than bar 2 despite the identical harmonic structure.

10.4.3 Listeners' and Written Responses

PS and PE

There is general agreement between the listeners for the positions of the PS, PE and areas of EOP, for both the MIDI and performance renditions and for written responses (the graphs are given in appendix 3.6.3). There are four areas of peaks of PS responses: bars 0.666 – 1, 2.333 – 2.666, 4.333 – 4.666, and 6.333 – 6.666. Combining the PS responses of these areas over 80% of participants choose bars 1, 2 and 4 and less than half choose 6. A similar proportion chooses the preceding positions as PEs. This supports the discussion above that even though exactly the same harmonic pattern is underpinning both bars 2 and 6, the melodic (and primarily rhythmic) characteristics override harmonic considerations for some listeners.

Overall, the graphs show that there is stability between listenings by the same listeners, though there are some differences. In more detail, in bar 2.333-2.666 in the first listening most of the responses occur on the beginning of 2.666 and on the semiquaver that precedes it. By the final listening, this position is being anticipated with responses in the first 3 notes 2.333 and most responses late in the beat.

In bar 4 the PS responses are spread more equally over 4.333 and 4.666. Again, most of the key presses of beat 4.333 are late in the beat and can be seen as anticipating 4.666. It is difficult to conclude where exactly the listeners decided the PS should be here as there is such an even spread over the two beats. However, the PE responses show that by the final listening most listeners lift their fingers during 4.333 indicating that the PE is felt clearly during that beat and not at the beginning of the next. This is supported by the results of the written responses and the click study (chapter 5).

In bar 6.333 most of the PS responses in the first listening are on the 3rd and 4th semi-quavers. By the final listening most are on the 2nd and 3rd semiquavers. Just like that of bar 4, this is a difficult start to synchronise because of its short duration, but some listeners learn and prepare for the exact position. However, the spread of responses is not just because of this difficulty. The graphic responses provided by musicians show very similar results (graph 3.6.3.11, appendix 3.6.3), even in showing spreads of responses. For example, there is a small spread of responses in bar 6.333 between the first and second semi-quavers. Almost all written responses give PEs and PSs on one note following the next. Only one musician marks an elided phrase (in bar 6).

The responses to the performances are on the whole very similar to those of the MIDI and to each other in that the same positions are identified, though there are differences in proportion. The responses for the Uchida performance are higher here than for Lipatti (in second session the responses for Lipatti are higher, chapters 3 and 4). The intensity contours are very similar. However, the tempo contours show that Uchida slows down more towards bar 6.333 that Lipatti. Overall, Lipatti's performance is a little faster than Uchida's. It therefore seems that, in a position where the possible phrase boundary is more 'hidden' by rhythmic features, one performer emphasises the division, while the other, emphasises the connection with the continuation and the listeners' responses coincide with this.

In general, the relative strength of positions described above (section 10.4.2) is supported by the listeners' responses (graphs 3.6.3.1-11 and chapter 4).

EOP

The relationship of the EOP responses to the PE and PS is systematic. For the first three MIDI PE/PS peaks, the EOP peaks are on the first beat of the bars that include the PE and PS i.e. on the arrival of the final chord of the phrase but not the final note, while the responses are a little more spread than for bars 7-8. Unlike for other pieces, there is almost no EOP response coinciding with the PE or PS.

The areas of EOP responses for the performances are the same but are more spread. The DL responses are mostly earlier than the rest and the responses for the later listenings are also earlier. All of these responses coincide with the beginning of the slowing down in the performance contours and the 'arrival' and then diminuendo. It may be that in the performances therefore, the hiatus and diminuendo have a strong influence on the position of the beginning of listeners' expectations.

Overall, the results of the different ability groupings are very similar (chapter 3). For the MIDI, the largest difference in decisions is in bar 6 where more DL and M than N respond. This may be because the DL and M expect the symmetry between bars 2 and 6 and, identify the similarity in the harmony and other musical features. These are confounded by other musical features at this position that do not emphasise this phrase boundary (such as the continuous semiquavers), and it seems that a larger proportion of the N than the other groups are relying on these contradictory features.

Summary of listeners' responses

The MIDI responses for PS, PE and EOP show more agreement between subjects and greater stability between attempts for each subject than in some of the other case-study pieces. The responses to MIDI and performances are similar though there are some differences. The three phrase parts (PS, PE and EOP) are more distinct here than in many of the case-study pieces allowing clear identification of the musical characteristics that relate to each one. Despite the few ambiguities discussed above, this extract is relatively straightforward in its phrase structure. However, not all of Mozart's music is as straightforward as this excerpt: there are often 'unpredictable elements, apparently in conflict with the prevailing continuity and regularity of style. He consistently introduces unexpected ideas, such as fresh melodic material, surprising harmonic changes, and irregular phrasing within a context of regular phrase structure' (Davis 1966, p. 25).

10.4.4 Performance contours

10.4.4.1 Tempo

Graph 3.6.3.7, appendix 3.6.3 shows that the tempo contours are relatively similar. That of Uchida is slowest overall and has the greatest note-lengthenings. The greatest tempo changes are lengthenings in bars 2, 4 and 7 by the performers, and bar 6 by Lipatti and Uchida. These coincide with the final notes of the phrases identified by listeners, annotators, and theorists (10.4.3 and chapter 7).

10.4.4.2 Dynamics

The intensity contours are also very similar (graphs 3.6.3.8-10, appendix 3.6.3). They all have three areas of minima, coinciding with the PE and EOP responses, in bars 2, 4 and 8. At each of these, the lowest trough coincides with the listeners' PE and is preceded by a relatively low minimum at the start of the bar. Between these minima there is a general rise and fall of intensity, two smaller ones for the first two phrases and a larger one for the last phrase (chapter 7).

There are some differences between the contours. The Uchida recording emphasises the gradual changes in volume more than Lipatti. The most noticeable difference is that there is a minimum in bar 6.333 (first semi-quaver) of the Uchida contour, but only a relatively weak one in Lipatti and Brendel contours. In the first session the listeners' PS/PE (and to a lesser extent the EOP) responses to the Uchida recording at this position is higher than that to the Lipatti recording. The listeners' response in the second session to the Lipatti is a little higher than the first session and that for the Uchida is a little lower than the first Uchida session, indicating an effect of the previous listening session. Like for the Bach Suite, although the difference between the responses to the different recordings as a whole is not statistically significant, local differences in responses to the different recordings coincide with local differences between the performance features in the different recordings (chapters 7 and 10.3).

Moreover, the Lipatti and Uchida tempo contours show that the tempo changes in the two performances are almost identical at 6.333 (graph 3.6.3.7, appendix 3.6.3). It seems, therefore, that the dynamic changes alone here have a considerable effect on the PE/PS and a small effect on the EOP responses at this position.

10.4.5. The features in view of listeners' and annotators' responses

10.4.5.1 Responses and the features

The feature graph 3.6.3.12 is presented in appendix 3.6.3. The Mozart Sonata has, in general terms, one of the clearest phrase divisions of the case-study pieces. Two main phrases (bars 1 – 4 and 4 – 8) and two further phrases (1 – 2, 2 – 4 and 4 – 6, 6 – 8) were identified. Each PE follows cadential and voice-leading preparations and has lengthening in the performances. The smaller phrase boundaries (2 and 6) also have pitch jumps and the bigger phrase boundary (bar 4) has a long note, an ornamented repeat of the opening, and is clearly in the home key. The responses show that listeners and annotators mark PSs in all four of the areas expected from the analysis (section 10.4.2), with a proportion marking bar 4 than the others.

The areas of the identified phrase parts are, like in the other case-study pieces, broad. This may be because of a number several musical characteristics including:

1) The piece begins on an upbeat. There is a spread of responses over the first three beats. Some listeners respond on the upbeat, while others do so on the first down beat; some marking the beginning of the phrase, and others marking the first strong beat in the phrase.

2) The PSs in bars 2 and 6 are on weak parts of the second (weak) beat. In bar 2, many respond on the nearest strong beat (bar 2.666). In bar 6 there is more ambiguity. Some identify the phrases as being elided (PE and PS on the first semiquaver of the 6.333). Others identify the PE on the first semiquaver of 6.333 and the PS after the leap up and on the second semiquaver of the beat, while still others wait for the stronger beginning of the upbeat (matching the earlier PSs). In bar 4 there is a similar spread of responses over the last crotchet and the beginning of the next bar.

The responses show that when presented with clear harmonic and melodic features, and inexact repetition of themes, listeners can make decisions as to phrase boundary areas, even if the exact position is difficult to identify or indicate.

Unlike for the Bach Suite (chapter 10.3), here the PEs are relatively clear to listeners. For each of the PEs all the features are located in the same place. There is general agreement among listeners and the PEs are more easily identified than the PSs though overall, the PS and PE responses seem rather equal, indicating equal importance of the two. This may relate to the clear PS and PE features (such as, repetition and cadence respectively for bars 4ff.). The PE and PS, for all but one of the phrases here are separate – the phrase ends on one note and begins on the next. For the PE/PS in bar 6 the distinction is less unanimous: for some the PE/PS occur on the same semiquaver, for others, the PE occurs on 6.333 and the PS on the next semiquaver. Each of the phrase boundaries is preceded by EOP responses, the first three peaking on the first arrival on last chord of the phrase and the last starting with the cadence.

There is very little difference between the responses to the different performances apart from those mentioned above (section 10.4.4.2). However, there are some differences between the responses to the MIDI recordings and to the performances. In bar 2 the majority of key presses in response to the performances occur on 2.666, while in response to the MIDI they are a little earlier - in the middle of 2.333. This could be because the performers accentuate the beginning of the third beat clarifying the arrival on the relatively stronger part of the PS, while for the MIDI, listeners respond more to the PE (and immediately the next PS). In bar 4 the trend of responses in the middle of 4.333 and the beginning of 4.666 is the same in response to all the performances. However, the response drops off in the MIDI in the following beats, while it continues to the end of the bar and the beginning of bar 5 in response to the performances. Again, this could be because the performers lead up to the main beat, this time the first beat of the bar, while in the MIDI there are no such performance features. In bar 6 the responses are much more similar, though even here the response on the third beat is higher in the performances than in the MIDI. The reasons may be the same as those suggested for bar 2.

These results may indicate that the phrase responses to the performances are being related to the metrical (beat) structure more than responses to the MIDI. The performance cues therefore may not only be clarifying phrase structure, but may be doing so in relation to low level beat structure and higher level metrical structure.

10.4.5.2 Summary of features according to bars

The most feature intensive areas are bars 4, 8, 2, 6 and 1. There is a distinction between the last note of one phrase and the first note of the next (section 10.4.5).

Bar 4

Bar 4.333 is the beginning of an area that follows the cadential progression and voice-leading. It is preceded by a long note which begins on a bar line. All of these features contribute to the 'feminine' ending. This is followed in 4.666 by inexact repeat, long note, four-bar template and change in texture. On bar 5's bar-line the fuller texture returns. This area, therefore, has both preparatory or predictive and confirmatory features. The PS, like bars 0.666 and 2.666 begins on an upbeat, with the next strong beat (strengthened by the fuller texture) occurring at the start of bar 5. There are some who suggest that a feminine ending (a resolution of a cadence on a weak part of the bar) is a weak ending and indicates a weak PE. However, this example shows that there are cases in which a feminine ending coincides with a strong PE.

Bar 2

Bar 2.333 is also the beginning of an area that follows a cadential progression and voice-leading (closing on an interrupted cadence, on a weak beat). During this beat there is also a pitch jump and a long note. A long note also precedes 2.333 as does a bar line. There is again a pitch jump onto 2.666, a boundary feature which is followed by the bar line on 3. As in bar 4, these features contribute to a feminine ending followed by an upbeat phrase. Again, a perfect cadence does not seem necessary for a PE to be identified

Bar 8

The first beat of bar 8 is the last note of the extract presented to listeners. It too follows a cadential progression (perfect cadence) and voice-leading, and some increase in the underlying rhythm early in the bar, which finally resolve on 8 onto the tonic in both melody and bass. It is also the first phrase to end on a bar line. All these features combine and together indicate the strongest PE so far. This is indicated by the performance features which have the slowest tempo and the highest intensity changing to the lowest in the extract.

Bar 6

Bar 6 has the same features as bar 2. The two notes preceding 2.333 are of equal length and longer than those that precede and follow them which, for Lerdahl and Jackendoff (1987) can also signal grouping boundaries (chapter 8).

Bar 1

Bar 1 also has two features: bar line and change in texture, which occur with the first strong beat of the piece and phrase. A small proportion of MIDI and performance listeners respond on this down beat but most identify the PS as occurring in the upbeat. This upbeat and the tendency for some to press on the downbeat, indicates an 'instability' of the beginning of this phrase, propelling us forward to the strong beat.

10.4.5.3 Major characteristics of phrasing features in the Mozart Sonata

The responses described above coincide with voice-leading, cadential progressions, and repetitions. There are other features throughout the piece (bar lines, long notes and a pitch jump) but these do not only coincide with large responses, indicating that these features, on their own do not seem to have the same effect on phrase perception. Once combined with other features, in this case, cadential progression and voice-leading and repetition, they do seem to make a difference between higher and lower responses, as can be seen for example in the difference in response and features between bars 2 and 6 with and without the long note. This indicates that it is not only the presence or absence of certain features that has an effect on phrase perception, but their combination with other features.

The results here, as in other pieces, also indicate that a full cadence in the local tonic does not seem to be a pre-requisite for phrase identification (Rothstein 1989). Any cadence (interrupted or imperfect, feminine or masculine) seems to be strong. This should be verified by using longer extracts to see if the shorter phrases are incorporated, and cease to be identified as independent, in larger contexts.

This discussion shows that even in such a 'simple' piece, where the performance contours are in general very similar, the differences between them influence listeners' responses and even affect later hearings of the same piece. Moreover, even such a piece reveals aspects of phrasing, including upbeats, clear separation between phrases and elisions, and the relationship with features, including voice-leading, cadences, repetition and long notes.

10.5 A relatively simple example, the strength of symmetry, repetition and contrast: an excerpt from the aria no. 25 'In quegli anni, in cui val poco', Act IV of *Le Nozze di Figaro* by Mozart (Mozart Aria)

10.5.1. Introduction
10.5.2.The piece
10.5.2.1 Text
10.5.2.2 Harmony, Melody and Change (or Contrast)
10.5.2.3 Vocal and orchestral parts
10.5.3. Listeners' Responses
10.5.4 Performance contours
10.5.4.1 Tempo
10.5.4.2 Intensity
10.5.5. The Features in view of the listeners' responses
10.5.6 Major characteristics of phrasing features in the Mozart Aria

10.5.1 Introduction

The extract from Aria no. 25 for Basilio from Act 4 of *Le Nozze di Figaro*, bars 107.5-123 (Mozart Aria) is unusual among the pieces in this study in that it is not taken from the beginning of a piece or movement.[90] Basilio's aria starts 105 bars earlier and the most recent double bar is 5 bars earlier. The aria finishes 18 bars after the end of this section. One problem with using a section from the middle of the movement is that, as listeners reported during the sessions, they tend to assume that the beginning of an excerpt is the beginning of the piece and therefore the beginning of a phrase. Here the beginning of the excerpt and therefore, for many listeners, the first phrase is chosen by the experimenter.

The excerpt is 17 bars long and the first half (bars 1 – 9) is repeated exactly (bars 9 – 17). Each half is itself made up of two distinct sections: loud F Major chords followed by a quieter melody in B*b*(the key of the movement).

10.5.2 The piece

10.5.2.1 The text

The four-line text by Da Ponte has two couplets. In the first, each line is divided by commas and the lines do not rhyme, In the second, the lines are undivided and rhyme. The setting of the two couplets seems to follow the text structure. In the first both the vocal line and accompaniment are divided and the rhythm of the notes mirrors the natural rhythm of the text. In the second couplet, the two lines are treated together, the peak of the vocal line pitch-wise, and the 'turning point'

[90] The Wagner is also from the middle of a movement, but the distinction between it and its surrounding sections is stronger.

harmonically (the beginning of the cadence), or the upbeat to it occur at the change between the lines. In the setting four-line text is repeated exactly, and is both times set to the same music.

10.5.2.2 Harmony, Melody and Change (or Contrast)

Bars 1 – 5 (and 9 – 13) are on the same chord. The orchestra and vocal part alternate. The orchestra gives the whole chord and repeats the same rhythm, and the voice answers with the same repeated octave, this time leaping down and varying the rhythm. The first time the vocal part does not have an upbeat, in the rest of the repetitions there are quaver upbeats. When this section is repeated, the first crotchet is replaced by a dotted-quaver, semi-quaver (to match the rhythm of the words). The pitches and harmony are therefore the same for these 5 bars when they are repeated but there are slight changes in the rhythm. The contour combined with implied metrical stress and the rests in between the outbursts, implies four divisions within this five bar section, matching the sections between the commas in the text.

The melody (bars 6 – 9), opening on B*b* and ending with a cadence in B*b*, is clearly in B*b*. It is only in bar 6 that it becomes clear that the f chords of bars 1-5 are the dominant of B*b*. In this rôle, they could be seen as preparing for the B*b* section. However, the two seem unconnected and the B*b* section seems incomplete (indeed after the second repetition (bars 14-17) the melody continues. Despite this, the complete change between bars 1 – 5 (and 9 – 13) and 6 – 9 (and 14 – 17) implies a clear division.

10.5.2.3 Vocal and orchestral parts

The contrasting relationship between the individual voices between 1 – 5 (and 9 – 13) and 6 – 9 (and 14 – 17) again implies a division between these sections. Even in this short extract there are different relationships between the voices. Bars 6 – 9 and 14 – 17 have a melody in the voice, horns and violins, and an accompaniment in the rest of the orchestra. In bars 1 – 5 and 9 – 13, there is alternation between the voice and the rest of the orchestra (which is playing together). The four sections are connected by a short passage in the horns and bassoons.

10.5.3 Listeners' Responses

Graphs 3.6.4.1-6, appendix 3.6.4 show the areas of high response and that they coincide in the MIDI and performance responses. In general there is also close overlap between PS, PE and EOP areas indicating that: 1) in some cases the areas of PE and PS overlap, and 2) in many cases there is relatively little preparation for the PE. Each of the areas of high response, and the possible musical reasons for these patterns of response, is discussed below.

Bars 5-6 and 13-14

As in the Bach Passion and the Bach Suite (chapters 10.2 and 10.3 above), the listeners' responses indicate a temporal gap between some PEs and PSs. This is especially clear in bars 5-6 (and their repeat in bars 13-14) in the MIDI responses where there is a spread of responses with peaks at 5.5 and 6. The PE responses peak at 5.25 (coinciding with the last f) and continue through the rest of the bar. The same pattern in seen in bar 13. As in the previous examples, the end of the previous phrase has arrived and there is an expectation for the next phrase to begin (the rest of bar 5 could indeed be seen as an upbeat to bar 6). The EOP responses also peak on 5.25 and 13.25, indicating that there is not much preparation for this PE. As well as the change in motive and texture and the rest in the melody line between the sections, the listeners expressed the impression that they knew the patterns in bars 1-5 could not continue forever. For some, the arrival of a change would signal a new phrase, though they did not know when to expect it.

There is also a spread of responses to the performances. In the Solti PS responses, there is a peak on bar 6 while in the Böhm PS responses there results are more spread with one peak starting from 5.4375 and another on 6.125 (and a similar pattern at bars 13-14). In the PE responses the Böhm again peaks earlier (on 5.25, coinciding with the last crotchet f) while the Solti has a high response at 5.4375 and a peak at 5.5 (the beginning of the rest). In the parallel position in bars 13-14 these trends are even more extreme. The Solti PE starts at 13.4375 and continues to 14.25 and the Böhm PE responses start at 13 and continue to 13.5.

The performance contours (section 10.5.4) indicate the reasons for the differences. Böhm's performance lengthens 4.25 and 4.75 while Solti's performance lengthens only 5.25. Similarly Böhm's performance lengthens 13.5, 14 and 14.25 while Solti's performance has only a very small lengthening on 13.5. In terms of the intensity contours, the Solti seems to lead up to the fourth repetition of bars 3-4 and 12-13 and then descend down to bars 9 and 17 treating the sections of the dominant chords (bars 1-5 and 9-13) as an upbeat to the following I-based bars. The Böhm recording has peaks at the same positions but additionally has peaks at bars 7.5-8.1 and 15.5-16 with troughs at bars 5, 9 and 13. Both the tempo and intensity contours indicate that the performance features of tempo and intensity change represent different interpretations of the same extract, with Böhm subdividing much more than Solti.[91] The differences between these characteristics of the performances are reflected in the responses. The Böhm EOP responses begin a beat earlier on 5 and 13 but for Solti, the response is in the same location as for the MIDI. These responses indicate that Böhm's

[91] Ostman's recording seems to be a combination of the two, with both the overall connection of the 8 bar sections of the Solti contour (clear in the intensity contour) with a subdivision 'on the way down' for bars 5-6 and 13-14 (see chapter 7).

performance prepares this position as a PE and PS, to which the listeners respond, while Solti's performance does not.

Bars 7-8 and 15-16

New text-lines begin on bars 7.75 and 15.75. For all renditions, there is a small response on 7.875 but almost no response in bar 15. It seems that the response at 7.875 is more of a continuation of the template than related to the specific features of the music there, especially as this position is abandoned by almost all listeners in the repetition.

Bars 9-10 and bar 17

There is a smaller spread of MIDI responses at bar 9. This coincides with a smaller temporal gap between the arrival of the resolution on bar 9 and the beginning of the new material on 9.5 (especially in the MIDI where the distinction between the 'solo' vocal line and the accompaniment is less clear). There are some PS responses at 9 but the peak is at 9.5. The PE peaks on 9 (and 17) as does the EOP, though there are responses before 9 for the EOP (especially at 8.5, 8.75) and 16.375. Here there is a little more preparation (the cadence, the descending melodic line, the longer note values) – enough to enable listeners to respond exactly with the last note of the phrase, but not enough to have an earlier EOP as in other pieces. This may be partly because of the fast tempo.

A similar pattern is seen in the performances as for the MIDI in bars 5-6 and 13-14. The PS responses peak on 9.5 for the Solti and 9 for the Böhm (and so anticipate). The PE responses peak on 9.5 and 9.75 for the Solti (and so are 'late' in comparison with the MIDI and the Böhm) and the Böhm peak on 9.0. The EOP responses in Böhm begin on 7.875 and continue until bar 9 while the Solti responses are lower but over a similar range. The Böhm tempo contour shows a slowing down to 9 while Solti lengthens 8.75 and 9.5. These are the lowest intensity areas for Solti. For Böhm the difference between these troughs and those for bars 5-6 and 13-14 is smaller.

Bars 1-4 and 9-13

There are also MIDI PS, PE and EOP responses at 2.25, 3.25 and 4.25 and PE responses at 10.25 and larger ones at 11.25 and 12.25. In bars 10 – 12 the PS are spread over the second and third beats (10.25 – 10.5, 11.25 – 11.5 and 12.25 – 12.5). So the orchestral PSs anticipated in the first half are now more accurately responded to. There are also responses to the performances around these positions but in these responses the accuracy is very similar in both halves. In general, the responses are higher in the Böhm than in the Solti (in the first session). In the Solti there seems to be greater distinction in the EOP between bars 1-4 and 5-6 (and 9-12 and 13-14). This may be because Böhm emphasises the smaller units more than Solti through changes in tempo and intensity.

Summary

In general, the responses show 1) two levels of popularity: the high responses to bars 1.5 and 9, and the lower responses to other areas and 2) different distances between EOP, PE and PS responses. The EOP responses are here particularly close to the PEs and PSs, suggesting little or no preparation and expectation. Even in such a short and 'simple' excerpt the results show differences in responses to different performances, which may be related to performance features.

10.5.4 Performance contours

10.5.4.1 Tempo contours

The tempo contours (graph 3.6.4.7, appendix 3.6.4) show that there is a very small range of tempo variation in these excerpts and that the various performances are very similar to one another. In many cases, the locations of the tempo variations coincide with the phrase parts identified in the MIDI responses.

10.5.4.2 Intensity contours

The intensity contours (graphs 3.6.4.8-10, appendix 3.6.4) show that the overall intensity ranges for the Böhm and Solti performances are very small and similar. The areas of minimum intensity that appear in the Solti also appear in Böhm contour. Overall, the Solti contour is smoother with more gradual and longer diminuendi. This seems to be reflected in the listeners' responses to the performances; the Böhm has more PS and PE responses than the Solti. The contour of the Böhm performance shows a peak at bars 6.5 – 7.25. In the responses, the EOP begins at 6.75 and peaks at 7.5. It may be that EOP begins after the peak of intensity (representing the peak of the phrase).

Unlike predictions that the same phrase is often being repeated in the same way (Friberg and Battel 2002, chapter 7), the minima in the Solti are more dramatic in the first half than in the second, and both the Solti and Böhm contours have a much larger breath at bar 4 than they do at the parallel place (bar 12).[92]

10.5.5 The Features in view of the listeners' responses

The above analysis of the piece and listeners' responses indicates that there are a number of musical features that listeners and performers respond to: contrast between the sections, harmonic preparation (and lack of it), and rests. Even in such a 'simple' and short piece there are a number of different possible interpretations of the phrase structure. The clearest phrase boundary is at bar 9 with a PE and a PS. As will be shown below, this is also the position with the

---

[92] The repetition in Ostman's contour is generally louder than the first half.

most features and a combination of all the different feature-types. Bar 17, being the end of the extract has one group of these features (the preparatory ones). Positions 5-6 and 13-14 are the next most popular choices, the next most strongly highlighted in some performances, and are also the next most-feature intensive areas. Bars 1-5 and 14-16 have a smaller number of features spread across them, are chosen by the smallest number of listeners, and are least accentuated in the performances. The features graph (3.6.4.11, appendix 3.6.4) shows the location of the different features identified in this study. The different positions of the piece are briefly discussed, beginning with the areas with most musical features.

Bars 9 and 17

The most feature intensive area is bar 9. As in the phrases of the Bach Passion, here there is a gap between the PE and PS. Bar 17 has the same features as bar 9 and is also included here. On 9 and 17 there are bar-line, following cadential progression, explicit voice-leading and increase in underlying rhythm, all of which took place during bar 8 and which resolve on the first beat of bar 9. On 9.5 there is, in addition, exact repeat, following rest, change in texture and change in motive, and four-bar template all of which, as discussed above, coincided with a high response both in the form of a PE primarily on 9 and a PS from 9.0 to 9.5.

Bars 6 and 14

The next feature intensive areas are bars 6 and 14 with bar-line, pitch-jump, rest, change in texture and change in motive. As discussed above, there is no preparation, but for some listeners there is an expectation that the repeating patterns of bars 1-5 will come to an end. However, for other listeners this was not considered a phrase boundary in the same way as the one at bar 9. This could be because not much happens thematically in these bars, and because the whole section turns out to be in the dominant of the theme and so acts as an upbeat to it.

Bars 1-5 and 9-13

These bars have inexact repeats, rests, pitch jumps and long notes, and alternating patterns between the solo and the orchestral parts. The responses indicate that some listeners respond to the octave leaps in the vocal line and the rests and repeats in both voice and orchestra, even though the resulting sections are very short and the vocal and orchestral parts overlap.

10.5.6 Major characteristics of phrasing features in the Mozart Aria

The above analysis of the features and listeners' responses indicates that the preparatory features (cadential progression, voice-leading, change in underlying rhythm) influence the expectation and identification of the PS, PE and EOP. However, if these features are not present, identification of PS and PE is possible (though less expected) as long as other confirmatory features (such as repetition, change in texture and motive) and boundary features (such as rests and pitch jumps) are present. The overlapping orchestral and vocal parts in bars 1-5 and 9-13 seemed to be less disturbing here than in the Bach Passion.

The text structure seems to be reflected to some extent in the music, though the second couplet is treated as one unit and not divided. For this piece, the text structure can help predict the possible musical phrases-structure but not automatically.

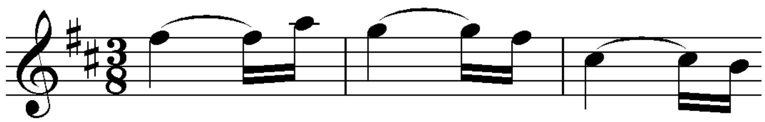
In general, the stronger boundaries seem to occur when the three different feature-types are present (preparatory, boundary and confirmatory). However, even if they are not all represented (such as in bars 1-5 and 9-13), phrases are still identified.

10.6. A clear yet unclear structure: Brahms' Intermezzo Op. 119, no. 1 (Brahms)

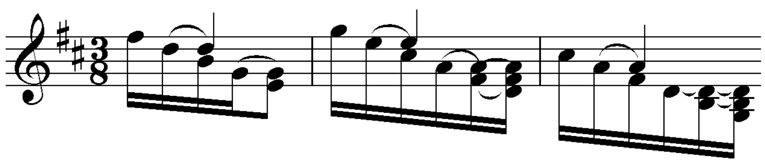
> 'Brahms' music is characterised by an avoidance of straightforward relationships. There may often be a simple aesthetic framework for his ideas, for example in periodic phrasing. At some level of the structure however, Brahms usually creates a functional ambiguity, giving his music its typically elaborate and complex character' (Dunsby, 1981, p. 1).

10.6.1 Introduction

On one hand, the opening of Brahms' Intermezzo Op. 119, no. 1 (Brahms) is a melody
Example 1:



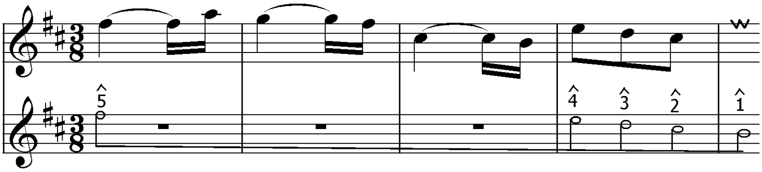
with an accompaniment.
Example 2:



On the other hand 'these bars are equally suggestive of a contrapuntal invention, where the horizontal structure of each part is more comprehensible than the background succession of vertical relationships' (Dunsby, 1981, p. 89).

On one hand, the opening three bars fulfil one condition of tonal definition (one of the characteristics of the opening section of any piece) by using exclusively all the notes of a key: that of D Major, with each note being repeated at least once. On the other hand, these notes also create B minor and the first three notes of the piece form the tonic triad of B minor.

On one hand, the two suspensions of the (antecedent) melody 'resolve' on the dominant and tonic notes. On the other, these resolutions arrive on the last semiquaver of each bar preparing for the first beat of the next bar (not necessarily reaching the expected harmonies). At the same time the melodic background begins a progression towards the 'tonic' b, which should be in bar 5 but is delayed (Dunsby, 1981, p. 91).
Example 3:



There are many aspects that are ambiguous in the first sixteen bars of the Brahms; contradictory information is given, expectations are set-up and then thwarted or the resolution is hidden. This all takes place within an outwardly regular phrase-structure where it is possible to present divisions as 'almost unambiguous' (Dunsby, 1981).

In this section the 'irregularities' that lead to the ambiguity are explored with the aim of identifying the reasons why, despite the ambiguities, there is "an outwardly regular phrase structure" that ultimately all the analysts that discuss the piece refer to and agree on, and that many of the participants of the study also identified. This will be done through a discussion of the harmonic and tonal ambiguities, the voice-leading, and the motivic and rhythmic characteristics of the opening (in relation to the whole piece). The piece will then be analysed from the perspective of the Gestalt principles discussed by Wertheimer. The listeners' and annotators' responses and the tempo and intensity changes of performers' will then be presented.

10.6.1 The piece

10.6.1.1 Harmony: The search for a tonic and the identification of the dominant

Returning to the tonal ambiguity of the opening, Dunsby suggests that there is hope: 'Brahms chooses to resolve the tonal ambiguity in b minor, altering the a's of bars 1 – 3 to a# in bar 4' (Dunsby, 1981, p. 92). The 'resolution' if it is felt as such, is obscured; the b is delayed and is only sounded, in the melody line, on the last semiquaver of the expected bar (bar 5) (Dunsby, 1981, p. 91). When the b does finally arrive, it is accompanied by a g in the bass, which is reiterated in the middle part. There is another b in the bar, on beat 2 in a middle part (with a G below it). The positioning of the b here highlights it, having parallels with the b in bar 1, thus increasing its importance. Bar 5, however, begins not only with the F# of the beginning but also with the D in the bass, highlighting the D tonality. Having been delayed from the first quaver, the b resolution not only comes after a delay and a subversion, it also arrives 'late' in relation to the metrical structure and after a new beginning has been felt.

Having arrived at the end of bar 5, as if in acknowledgement of the difficulty, an alternative 'resolution' (Dunsby, 1981, p. 92) follows in bars 6 – 7. Here, the a is kept natural and D is expected, which arrives on bar 7.333. After the 'resolution' the a# returns at the end of the bar taking us back to B. It is as if bars 6 – 7 could replace bars 4 – 5. Two options are presented to tackle the problem but in neither provides a satisfactory resolution. If bars 6 – 7 really had replaced bars 4 – 5 there would be increased support for D major and the phrase length would be reduced to 6 bars. As Schoenberg says in a later example of an eight bar phrase 'if 8 measures constitute an aesthetic principle, it is preserved here in spite of the great freedom of construction' (1975).[93]

Within the two bar groups of the opening, expectations are generated. At the end of bar 4 we expect a resolution to b, which is frustrated and replaced by an attempt at D. This does not last long, with another veer back in the direction of b at the end of bar 5. This again is contradicted in bar 6 where the a♮ returns with a clear A^7 chord, (preceded by D) and leading to an expected D though we have to wait for the root until the second beat (Db-a). The next bar, bar 8 outlines F# (and its dominant C#) directing us back to B and leading back to the beginning and B (in-as-much as the beginning was in B).

Bars 9 – 10 are exact repeats of the opening while the first half of bar 11 is modified and suggests that there will finally be a resolution of the ambiguity between D and B by sharpening the a♮ of bar 3 to a# (indeed Hinson in his edition of the piece makes the relationship explicit by connecting the stems of the a# to the b at the end of the bar, a connection he does not make anywhere else in this opening Hinson, p. 112). This does not last long however, the a# being flattened again an octave lower at the end of the same bar. Though we get a repeat of, what Dunsby called the 'resolution' to the tonic b in bar 3 again in bar 11, bar 4 is modified in bar 12. The melody is transposed down a tone so the line no

[93] At the end of the piece (bars 47 –54) the first presentation of the material is exactly the same as the opening and the second presentation gives the b in the melody voice in the fourth bar (bar 58), where it 'should' have been to resolve the descent (example 3). This time, it is even anticipated in the melody at the end of the previous bar. Here again the 'supporting' harmony does not reinforce b as a 'home' tonality. Only in the penultimate bar (its second beat, because of a suspension) does the resolution to B arrive.

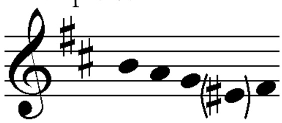
longer leads to a potential strong b in a direct scale on the first beat of bar 13 (equivalent to bars 4 – 5, see example above). The melodic line is changed to:
Example 4:

bars 4 and 6 bar 12

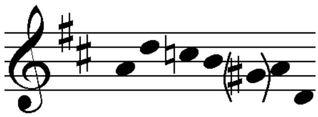
missing out the b that should have been at the end of the bar and accentuating the a by delaying it to the last semiquaver (a position of ‘resolution’ earlier in the piece). The get expected b arrives, delayed again to the second beat of bar 13 but sounded against an a in the bass.

So far then, the melody part has had several descents aiming for B, none of which have been satisfactorily resolved until bar 13 where at least the b is heard clearly in the melody part. However, even there is not a clear resolution. The section reached is described by Dunsby as having ‘strong chromaticism in substitute harmonies’ (especially bars 13-5) (Dunsby, 1981, p. 88).[94] These chromaticisms, though reached without a clear traditional modulation (which is not surprising in this concise form and ambiguous piece) are those of f# minor, dominant minor of B, which at the end briefly becomes F# major, the dominant of B minor.

The voice-leading in this section coincides with the underlying harmony in some ways more than earlier in the piece. The b on bar 14 sets off another descent,
Example 5:



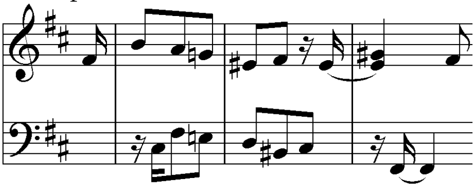
resolving, with its own leading note, on f# on the second beat. The bass line outlines a descent (jumping off from the dominant of f#) in bars 14-15.
Example 6:



These descents are not combined in a manner in which they reinforce each other. A canon (a device used for the first time in bar 4), which brings the bass in a beat after the melody, means that the resolution of the melody line is reached while the bass is still reinforcing the dominant note of F# with its own leading note.

[94] The ‘chromaticism’ here, begins already in bar 12 and continues to bar 16 and involves the addition of G#, E#, D#, (the flattening of E# and D# back down to E and D when there is a downward motion), and the recurrence of a# at the end of bar 16.

Example 7:



By the time the bass arrives on its F# in bar 16, the melody is preparing another attempt at a 'resolution' (an extended version of that of bar 12.666-13.333). Finally at the end of the bar (not the second beat which is where it should come if the rhythmic motive would have continued) we hear the F# chord we have been waiting for, though it is presented in the major.[95] The bass note of bar 16, F#^{1}, is lowest in the piece so far. It is as if the first note of the piece and the last of the section create a frame – possibly a dominant frame – for the section.

The expectations in the previous phrases were for the phrase end to arrive at the beginning of the bar. Here, the first 'clear' cadence comes at the end of a bar. The complete change of bar 17 indicates a new, and to an extent unrelated, beginning. There is little sense of the new section 'growing out of the previous one'. This is different from the situation of previous phrases and other compositions by Brahms where there is an overlap between the phrase end and start. Despite the reinforcement of the dominant of B for these five bars, bar 17 brings a clear D Major (reinforced by the c#s at the end of bar 16).

There seems to be a sense in which the tonality in the opening section of the Brahms, if it is defined at all, is defined through implication, through generating expectations of B. The dominant of B, both the note and the harmony, dominates the opening 16 bars, and indeed the piece as a whole.[96] As Dunsby notes, 'F# is a significant note at its position at the head of each phrase (bb. 1, 9, 17, 31, 47, 55, and for smaller divisions, bb. 5, 21, 43, 51 and 62)' (Dunsby, 1981, p. 103).

The importance of f# is supported by the 'unfinished descents' of bars 4 and 6 (example 3) and by the modified version of those two bars in bar 12. Because the descents of bars 4 and 6 are not resolved we are left with open c#s. The c# is picked up again, and most convincingly resolved to the b on bar 13.333. By this time though, as discussed above, the tonality of f# is dominating and, having had a hiatus, the scale continues down to f#. All against the background of f#.

Example 8:

[95] It is not that unusual end a minor section on a tierce de picardi and here is especially necessary if F# is seen as the dominant of B.

[96] a natural is heard in the melody five times (end of bars 1, 7, 9, 12 and 14). D strongly expected only once and this is in the 'alternative' bars 6-7.

10.6.2.2 Thirds as a structural force and as a surface feature

Alongside the search for the tonic, there is another related organising principle; that of a cycle of descending thirds both in the foreground and the background (Dunsby, 1981; Newbould, 1977, a more detailed discussion of these is given in appendix 10.6, section 10.6.1). Built on different intervallic and harmonic relationships, Dunsby's and Newbould's analyses both reach structural groupings and contradictions. The search for a tonic and the importance of thirds and fifths in the Brahms shows that the opening avoids establishing a tonality through the obscurity of its harmonic function. Dunsby sees this as the 'greatest innovation of the piece' (1981, p. 88). The acute harmonic ambiguity prohibits the definition of a key. In some cases (such as the String Quartet Op. 51, no. 1) the difficulty is because remote harmony pulls away from a tonal centre (Dunsby, 1981, p. 91). However, in this case, there is very little chance for the home key to be established before it is 'pulled' away from. The ambiguities here are not of the successive variety (discussed by Schoenberg, 1975) but of simultaneous types of organisation (Dunsby, 1981, p. 100). Rather than relying on ambiguity of phrase and proportion, Brahms' music is contrapuntally ambiguous, with different levels of organisation in vertical conflict. This produces tense internal articulation but, according to Dunsby, the principle of regular phrasing is often maintained, as in the case of this Intermezzo. This discussion shows that that it is partly maintained by generating expectations through voice-leading and harmonic suggestions which help to suggest that a phrase is likely to end and when. These expectations are also generated through other features including rhythmic and melodic factors which are here discussed in the context of three 'motives' that dominate the opening.

10.6.2.3 The 'Motives'

Both Dunsby and Newbould identify a 3 + 1 + 2 + 2 grouping of the first 8 bars and group the first 8 bars separately from what follows. This is partly because of the melodic, rhythmic and textural features of the first 8 bars, which can be described through the identification of two basic motives and their variants in the first 16 bars:

Motive a (seen first in bar 1 and immediately repeated in varied form) consists of the 'melodic idea' (Dunsby, 1981, p. 89) (example 1) and the 'main thematic cell' (example 2) (MacDonald, 1990, p. 360) of the descending thirds and no real bass. Motive a^1 consists of a variant of the melodic idea of motive a, using the same rhythm as the original while the interval is increased from a 3^{rd} (or 2^{nd}) to a 5^{th} (such as in bar 4). The 'accompaniment' is now a scale of 'vertical' thirds and a separate bass part has been introduced.

Motive b consists of a descending scale preceded by an anacrusis (bar 3.833-4.666). A 'new' rhythm has been introduced: quavers preceded by a semiquaver have replaced the crotchet tied to a semiquaver, semiquaver pattern. The connection of the semiquaver to motive b is clarified by the canon in the bass.

The middle part is 'verticalised' thirds and fourths. The bass part follows the top part in a canon beginning one beat after the top part and motive b^1 is an extended and varied b. Bar 16 combines the motive from b (this time only in the left hand) with the top part of motive a^1 (this motivic structure is summarised in column 2 in table 10.6.2.6).

Brahms avoids exact repetition and 'repeated phrases, motives and other structural ingredients only in varied forms, if possible, in the form of developing variation' (Schoenberg, in Frisch, 1981-2, p. 216). In this case, only two bars are repeated exactly (bars 1–2 in 9-10).

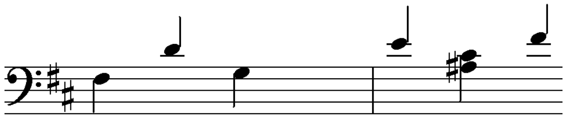
Having categorised the first 16 bars in this way, is there any way in which the reasons for the clear 'phrase' structure that the analysts mention can be identified?

10.6.2.4 Rhythm and Metre

Despite the two-chords per bar rhythm identified by Dunsby and Newbould, there is also a sense in which a rhythm of 1 unit per bar can be identified in the first three bars (motive a). A new chain of thirds begins at the start of every bar; over the bar-line of bars 1–2 there is a change of direction of the thirds (with octave displacement) (e – g^1) and over the bar line of bars 2–3 there is a 7^{th} (or a 2^{nd} with octave displacement). In this way the beginning of each of the bars is highlighted. The higher octave is returned to at the beginning of each bar and coincides with the beginning of each repetition of the 'melodic idea'. The final semiquaver has the dual function of finishing the 'melodic idea' and even 'resolving' it in bars 2 and 3 and at the same time, being the upbeat to the next. This is made explicit at the end of bar 3 where motive a overlaps with motive b (and its canon in the bass).

The 'rhythm' of motive b on the other hand, is three times as fast, with clearer harmonic changes on each quaver of the bar in the melody and bass, reinforced by the syncopated middle parts. This increased harmonic rhythm is expected to lead to a cadence and the end of a phrase or section. As discussed above, however, the expectation is not completely satisfied and instead motive a^1 arrives, restating the opening f#^{1}. As discussed above, there is then another attempt at a close, which again fails and leads to motive a^1. In this presentation of motive a^1, the bass fills in the final quaver of the bar, superficially increasing the rhythm to three again. At the same time however, the bass gives two rising 6^{ths} followed by a rising 4^{th} (avoiding an a# in the bass and still following through a rise to f# in the 'off-beat' quavers) (example 21a). This results in what can be seen as a weak hemiola effect so characteristic of Brahms' pieces. However, as it is not supported by the upper parts, this does not dominate.

Example 9:



When motive b returns in bar 12 it is shortened by a quaver allowing the canon to come through more clearly and allowing it to be repeated each time starting with a semiquaver upbeat to a new bar. This is especially the case in the upbeat to bars 12 and 13. This also has the effect of resembling motives a and a[1] which featured the lone semiquaver in the melody line at the end of the bar. By stressing the first note of the next bar with an anacrusis the rhythm is slowed back to one per bar. In bars 14 – 15 the two types of rhythm are played against each other (three clear units in the bar against one important one at the beginning) and, combined with the general harmony of the two bars (the relative clarity) and the voice-leading, the rhythm seems to increase back to three in a bar (this rhythmic structure of units per bar is summarised in column 3 in table 10.6.2.6).

If an increase in rhythm (units per bar) increases the expectation of the end of a phrase and a following return confirms the new start, then in these terms, as well as in the harmonic terms outlined above, phrase ends may be expected in bars 3,5-6, 7-8 and 11-15, and bars 5, (7), 9 and 16 may be seen as their, at least partial, resolutions.

The preparations for resolution (the melodic descent of the melody and the harmonic implications) have been discussed in relation to the preparation of bar 5 and the 7. The phrase boundary actually described by Dunsby, however, is not preceded by the closing motive b. Instead, in bars 7-8 there is a weak hemiola effect in the bass and motive a[1], the melody of which picks up on descents that we have seen before and this time 'resolves' them both.

The descent of bars 1 – 2 (example 1), returns in the melody of motive a[1], in bars 7-8 leading down from a[1] to f#[1]:
Example 10:

The intervening descent of motive b of bars 4 and 6 is outlined and to some extent resolved on the second beat of bar 9. In these ways some sense of expectation of the end of a phrase (and a stronger sense than in bars 1 –4) is generated, reinforced and, to some extent, borne out with the return of the beginning in bar 9. The return of the beginning is interpreted as implying that the previous phrase had finished and a different phrase has begun. Here the repetition of previous music associated with a 'beginning' is important. 'New music is never beautiful on first acquaintance … The reason is simply this: one can only like what one remembers, and with all new music that is very difficult … The great popular composers constructed their melodies by 'repeating every little phrase often enough for it to impress itself on the listener' (Schoenberg, 1931, quoted in Frisch 1981-2). Not only does the phrase 'impress itself on the listener', but here, in identifying a beginning that heard before, it is interpreted as a beginning here too.

However, the division between end and beginning is not clearly over the bar line with one phrase ending at the end of bar 8 and the next beginning at the beginning of bar 9. As has been implied, often the end of a phrase is arrives at the end of a bar bringing with it an expectation for the new phrase to begin at the beginning of a bar. However, here the 'resolution' can be interpreted as being delayed or lasting until the second beat of the bar and as a result, the end overlaps with the beginning. It is partly this type of overlapping that is contributes to the 'evolutionary process in Brahms' last piano pieces, especially where the thematic matter advances not by formal 'repetition but by one phrase growing naturally out of another' (Burnett, 1972, p. 169).

Additional evidence is given in bar 16 where the melody in the top part, when seen in conjunction with the semiquavers below it, is a modified version of motive a^1. This is the same motive of bars 7 – 8 in ending the first main phrase and so maybe has an 'ending' function now. This time however, instead of ending on a semi-quaver (that could be an anacrusis) and jumping down a fifth, it comes to rest on a quaver f#.

At the same time, the canon in the bass has increased importance. It could suggest that the "bar line" falls a beat later than written. This would mean that the last beat before the double bar, the beat where all voices have finally resolved becomes the 'second' beat of a bar. As has a been mentioned, the second beat seems to have particular importance as the position closest to resolutions, and finally a clear resolution arrives on it. This can be seen as an example of Brahms' style of 'developing variation' (Schoenberg, 1975).[97] Here a concurrent different position of the bar line is suggested by the bass and reinforced by the result in bar 16. All these point to an end of a phrase at the end of the bar.

On the other hand F# (major) is the dominant of B and in theory we should expect a continuation (and further resolution). Moreover, the previous phrase boundary (bar 9), and the 'smaller division' (bar 5, Dunsby, 1981), overlapped with the beginning of the next and ended (or, to an extent, resolved) on the second beat. There are some analysts (such as Koch, 1787, 1983, chapter 9) that stress the importance of beginning and ending phrases in the same place. This may contribute to our expectation of what will happen.

---

[97] 'In Brahms motivic development becomes more intense and pervasive. It permeates all parts of the texture and even brings to break down or obscure the phrase structure, as in the ambiguity created at the juncture between the halves of the sentence. Motivic development can also affect the metrical framework' (Frisch, 1981-2, pp. 225-6) as in the hemiola of bars 7 – 8 mentioned above.

In bar 17, what actually arrives is a complete change in texture, theme, (rhythm), a lower bass note than we have had so far, which is a D. A new section (and a new phrase) has begun not completely resolving what came before.

10.6.2.5 Applying the Gestalt principles

As was done in chapter 8, the case-study pieces can also be analysed using the Gestalt principles. In chapter 8 this was done on the basis of previous approaches and considered only some of the principles. For a more general comparison, all the principles were applied to the Brahms following Wertheimer (1924, see appendix 10.6 section 10.6.2). This showed that the more specific Gestalt Factors that are intuitively applicable to music, seem to be effective on the motivic level and in some cases can give more general terms for the identification of the phrase level in this piece. The general nature of the terms used in the gestalt principles means that it seems that they can be moulded to have the most useful implication for each situation.

10.6.2.6 Summary of the analyses of the Brahms

This short analysis has not covered all the elements of this opening. However, amongst all the ambiguity and contradictions of the piece, the analysis suggests that there are elements that elicit expectations of phrase ends (voice-leading and harmony, ‘rhythm’) and others that suggest that a new phrase has started (repetition of material or changes). The phrase boundaries themselves are not straightforward; with phrase ends overlapping with phrase starts. It seems that the expectations generated by harmony and voice-leading, ‘rhythm’, texture and motive combine to produce the ‘phrase’ framework and expectations. These expectations are both generated in this piece and result from associations with other pieces in the same genre. For example, phrases ends are expected to need resolutions (the strongest type of resolution being on the tonic), phrases may be equal in length (Koch, 1787, 1983), start and end in the same position in the bar (Koch, 1787, 1983), and are an even number of bars in length. Table 10.6.2.6 shows possible interpretations of phrase structure in column 4 along with the motivic and rhythmic information (columns 2 and 3), and structures suggested by the different analysts discussed above. To give the context of the whole piece, a more general representation is used for its later sections.

Table 10.6.2.6 Outline of structure					
A section: B minor /D major, Highlighting F#					
Bar no.	Motives	Rhythm	Possible PS & PE & *EOP in italics*	Beginning of Ground (Newbould)	F# beginning phrase, and smaller divisions (Dunsby)
1	a	1	1	1	1
2	a	1			
3	a	1			
4	b	3	*4*		
5	a^1	1	(5)	5	(5)
6	b	3	*6*		
7	a^1	1			
8	a^1	1			
9	a	1	9	9	9
10	a	1			
11	a	1			
12	b^1	1/3			
13		1/3	(13)		
14	b^1	1/3	*14*		
15		1/3			
16	a^1/b	1	16		
B section (bars 17 – 46): D Major			17 – (20-21) – 31	18	17, (21)
			31 – 42	32, 35, 38	31
			43 – 47	43	(43)
A section (bars 47-67): more clearly in B minor by the end			47 – (51) – 55	47, 51	47, (51)
			55 – (62) - 67	55, 59, 63	55, (62)

The following section discusses the listeners and annotators' responses, and performance contours in relation to these structures.

10.6.3 Listeners' and Written responses

Graphs 3.6.5.1-6 (appendix 3.6.5) show the written responses and listeners' responses to MIDI and performances for PS, PE and EOP. In general, the graphs show that although there are many responses throughout the piece, there are a number of positions favoured by listeners over the rest. There are four areas of high PS response (bars 1, 5, 7 and 9), with smaller peaks at bars 2.666-3, 3.666-4, 5.666-6, 8, 10.666-11, 12.666-13 and 13.666-14. These are all preceded by, or coincide with peaks in PE and EOP.

The relative importance of bars 1, 5 and 9 is not surprising in light of the above analysis. The next in importance in the listeners' responses is bar 7 which was described as an 'alternative' to bar 5 above. It is preceded by motive b, reiterates the highest note of the piece so far, and is preceded by the largest interval in the melody.

One third of listeners respond to bars 3, 11 and 13 for the MIDI, more do so for the first listening of the Kovacevich, while fewer do so for the Gould performances. Bars 3 and 11 are the same (except for the a which is now sharpened to the a#). The 'large pitch interval' could be contributing to the separation of what has come before from the beginning of bars 3 and 11 which are mainly chosen by M and N. Bar 13 is the equivalent position to that of bar 5, it brings the final resolution of the descent to the b and is the end of the first presentation of the sequence that dominates the last five bars of the section.

In general, listeners' responses show a concentration of PSs around the bar lines. This could lead to the suggestion that listeners are responding purely to the metrical structure. Metrical structure plays a part, especially in this piece where the motives usually fall within a bar (sections 10.6.2.3-4). However, if it was simply a matter of identifying bar lines (or hypermetrical levels) there would be an expectation for greater accuracy in key pressing especially as the music was presented in MIDI and so the bars were of equal length. Here, however a range of notes around the bar lines are chosen as PS and PEs.

Graphs 3.6.5.1-6 appendix 3.6 also show that although, the same positions are chosen by at least some listeners in response to all performances, the proportions of listeners' who respond at each position, differ between renditions (MIDI and both performances) and between sessions. In general, the three main PS positions are chosen by the majority of listeners in response to all renditions, though even here, the response is highest in response to the first session of the Kovacevich. Similarly, the response to the rest of the positions is also generally highest in the first session of the Kovacevich, with the Gould lowest, and the MIDI in between. The differences between the responses to the different performances are statistically significant (chapter 4). In the second session of the performances (i.e. when the listeners heard the second recording two weeks later), the responses to the Kovacevich in general fall, whilst those for the Gould rise and the difference between the responses to the two performances is no longer statistically significant (chapter 4). The differences in responses to these performances are amongst the most dramatic amongst the case-study pieces. As will be discussed below (section 10.6.4), the differences between the performances are also more dramatic than for the other case-study pieces.

Looking again at the written responses (graph 3.6.5.11, appendix 3.6.5), as in the listening studies, there seem to be two groups of interpretation: those that identify almost every bar and those that identify only positions 5, 7 and 9 and possibly bar 13. Also like in the listening experiments, there seem several options as to a location of a possible phrase boundary in the second half of the excerpt. As in the listening study, the ends of bars 12 and 13 following the rests are the most popular phrase start positions in the second half.

These results indicate that 1) musical features coincide with positions of responses, 2) performance features coincide with the proportion of responses and 3) there seems to be an influence of one performance on the second (section 10.6.4 and chapter 4).

'Learning'

Graph 3.6.5.6, appendix 3.6.5 shows the listeners' MIDI PS responses according to the three different listenings and shows that, for the positions with a high level of response, the largest change in choice of position between the first and final listening is that for bar 9. On first listening it was chosen by the same number of listeners as bars 3 and 13. By the last listening it has the third largest group (including the beginning of the piece) and is the same number as responded within bar 5. The EOP responses show that though the peak of responses is lower for the final listening than the first, it begins earlier. The PE responses show that listeners are lifting earlier with the later listenings.

The PS, PE and EOP responses suggest that the first time the piece is heard there is little preparation for the PE and PS in bar 9. However, by the final listening, the new phrase start is more expected and identified. If there were a strong expectation for the end (and therefore beginning), then there would be more listeners choosing bar 9 at the first attempt. If on the other hand, there were a very strong sense of continuation as opposed to end that could be overridden by the return of the opening, there would be less listeners choosing bar 9 as a beginning. It seems therefore, that repetition of opening material is more important than anything that comes before this phrase boundary that could lead either to the expectation of a phrase boundary earlier or lack of it here. There is also a slight increase in the number of listeners choosing bar 7 and those choosing the positions in bars 13 and 15.

There is an overall increase in responses over the three listenings. Some the decisions made in the last hearing are only possible it seems, after some familiarisation with the piece. In general, the differences are less dramatic for the performances.

'Experience'

Graph 3.6.5.5, appendix 3.6.5 shows the PS responses for three groups of musical experience. It shows some differences in the proportions of responses at the different PS locations; a larger proportion of DL choose bar 9 while Ns dominate bars 2-4 and a larger proportion of M and DL than N chose bar 5. The difference between the responses of the three groups of listeners is more extreme in the

Gould I and less so in the Kovacevich I. The responses to the Gould by the N indicate that performance does not always 'clarify' phrase structure for all listeners.

10.6.4 Performance contours

10.6.4.1 Tempo contours

Graph 3.6.5.7, appendix 3.6.5 shows the tempo contours. Gould's is the fastest with what seems to be a systematic relationship between phrase and hypermetrical structure and the slowest is Kovacevich's.[98] Having been composed in the 'Romantic era', this is the piece out of the 6 case-study pieces in which 'phrase final lengthening' (discussed in, for example, Friberg and Battel, 2002) is most likely. All performers lengthen the beat before bar 9 preparing the return of the opening theme and a phrase start. Similarly, all performers lengthen 16.666. However, these are the only two beats on which all performers 'agree'. This coincides with the clearest PS/PEs identified by listeners to the MIDI as well as the performances.

Gould, lengthens the first note of bars 1, 3, (5), (7) and 11. He also lengthens the first beat and, more so, the second beat of bar 12, the final beat of bar 13, and the second and third of 15 and 16. Compared to the results of the listeners to MIDI, Gould is emphasising the first beat of most phrase boundaries identified. However, the largest tempo changes occur on the upbeats to bar 9 and on the second and third beats of bars 15 and 16. These latter positions are clarifying the PEs whereas the other positions mostly coincide with the identified PSs.

For the group of positions with smaller tempo changes (bars 3, 5, 7 and 11), all are the first beats of bars and are approximately the same in length. Indeed, the first beat of bar 9 is only a little longer than these. The difference in importance, therefore, is temporally shown by lengthening the upbeat rather than the downbeat. The first beats of bars 3 and 11 (parallel positions) are not prepared but are simply lengthened. Those of bars 5 and 7, however, are preceded by the lengthening of the previous beats. Bar 5 may be seen as a PS and the repeat of the closing figure in bar 6 leads to the expectation of a PS in bar 7 (section 10.6.3). Gould emphasises both but then lengthens the upbeat to bar 9 much more, showing where the first big phrase boundary will be.

For Kovacevich, the greatest lengthenings are also on the upbeat to bar 9 and bar 17. However, this is where the similarities end. Unlike Gould, the lengthening at the end of the extract is more gradual. The next two most substantial lengthenings occur on the upbeats to bars 6 and 14, the next group of positions of the similar lengths are on the upbeats to bars 2, 7, 10, 12, and the next are the upbeats to bars 3, (4), 5 and 11. The greatest changes in tempo therefore coincide with the clearest

[98] The average tempo of Lupu's performance is between the other two and the tempo variation is generally over a smaller range.

identified phrase boundaries. The other tempo changes occur on upbeats to all the bars (except bar 8 which precedes the phrase boundary of bar 9). Changes in tempo, in these cases, coincide with the metrical structure and seem to be dictated by this rather than phrase structure.

10.6.4.2 Intensity contours

The intensity contours are similar (graphs 3.6.5.8-10, appendix 3.6.5). They gradually rise to the area of bar 8 and then fall, with a second smaller rise around bar 13 and the general positions of least intensity are similar. Indeed, the intensity contours seem more similar to each other than the tempo contours.

All the studied performance contours have the largest decrease in intensity over a short time during the last notes of bar 8. For Kovacevich and Gould there is also a minimum on bar 9.333. However, there seems to be a difference between the minima at these two positions: the one at the end of bar 8 is the end of a longer diminuendo, while the second results from a 'lengthening' of the accentuated first notes of bar 9 which, on the piano, do not ring out but die away, creating another minimum. Table 3.6.5.7, appendix 3.6 shows a summary of the positions of minimum intensity for the three performances. There seems to be a relation between position of the dynamic change, its size and the listeners' responses to these recordings.

In the Kovacevich, there is a diminuendo to bar 5, accentuation (i.e. relatively loud note) of the last beat of bar 6 leading to bar 7, and a larger diminuendo to bar 9. Bars 5 and 9 have a higher response than 7 but all three have high listener responses. It seems, therefore, that changes in both directions; accentuation through increase in intensity, and diminuendo, prepare listeners for PEs. The preparation for bar 9 through the longest diminuendo, coincides with a peak in listeners' responses on the last beat of bar 8, indicating that the end is expected there earlier than the equivalent position in bars 4 to 5, which is partly encouraged by the longer diminuendo and minimum.

The high intensity areas coincide with the main PSs identified in listener responses. It seems that these are more important than the minima except around bar 9. This is not surprising as in this piece, the clearest signal that there was an end of phrase is the next phrase start.

10.6.5 Musical features and listeners' and written responses

10.6.5.1 Summary of the features

The features graph (3.6.5.12, appendix 3.6) shows the feature combination. The most feature intensive areas, with 9 features, occur on the first beats of bars 5, 7 and 9 and somewhat lower number is on the first beat of bars 6, 17, 4, 13, 3, 8, 12 and 16 and 8[3]. The most regular feature is the bar line (of which there are 17) 8 of which occur with 1 or 2 other features. Only 1 bar line (bar 10) occurs alone.

All three of the most feature intensive areas have 7 common features: bar line, following a cadential progression, pitch jumps, implicit voice-leading, change in texture, change in motive, following increase in underlying rhythm. Bars 5 and 7 also share following explicit voice-leading, while bars 5 and 9 also share 'begin at the start of a four bar template'. The only feature that is unique to one position among the three (and among all positions in the piece) is the exact repetition of bar 9. There are two features that do not occur on the first beat of any of these bars – the beginning of imitation in the lower part, and following a long note or rest. Despite the seemingly small differences between the features of bars 5, 7 and 9, each position has a different character. Bar 9 is the main phrase boundary, bar 5 can be seen as a sub-phrase boundary and the bars before bar 7, as discussed above, encourage an expectation of a new phrase boundary which is delayed (or is an anticipation of) bar 9.

The next most feature intensive area is bar 6 which shares 4 features with bars 5, 7 and 9: change in motive, change in texture, pitch jumps, and bar lines. The two new features are inexact repeats and imitation in a lower part.

10.6.5.2 Features and responses

10.6.5.2.1 The three peaks

As discussed above (section 10.6.3), the most prominent response is the first beat of the piece. There are three distinct peaks in listeners' responses for all renditions (which are also peaks in the written responses), which coincide with the feature intensive areas on bars 5, 7 and 9 (section 10.6.3). However, their popularity amongst listeners is unequal. For all three performances the proportion of key presses follow the same trend: the most prominent is bar 9, followed by bars 5 and 7 with the responses to Gould for each position being most different. The differences in feature combinations at these three positions may explain the difference in response.[99]

The proportion of responses during the first semiquaver is similar for all three positions. Having heard the lead up to and only the first semiquaver of the bars, a similar number of listeners indicate a PS. During the second semiquaver of each bar the proportions of PSs changes. On hearing the third semiquaver of the bar there are further responses. For bar 7, the majority of listeners indicate the PS within the first semiquaver – they are expecting a PS here and indicate a PE and PS within the first quaver of the bar. There are few PS indications throughout the rest of the bar. In bar 9, a similar (slightly larger) number of listeners also indicate a PS in the first quaver. However, having heard the next semiquaver, more

[99] Though the pattern of responses is different in the written responses.

listeners are persuaded of the new PS and indicate it then. This difference between the two contributes to the possible conclusion that in bar 7, a new phrase is expected, while in bar 9 it is confirmed. A similar pattern is seen in the comparison of bar 5 with 9. There are more PSs on the second semiquaver of bar 9 than on that of bar 5, while there is a similar number of PSs on the first.

The next most feature intensive area is bar 6 which shares 4 of its 6 features with bars 5, 7 and 9: change in motive, change in texture, pitch jumps, and bar lines. The two new features are inexact repeats and imitation in a lower part. The responses on this position, however, are much lower than those of bars 5, 7 and 9. Instead, the next prominent responses are the first beats bars 3, 11 and 13. There are several positions with responses from 20% or less of the listeners, some of which coincide with a small number of features.

10.6.5.2.2 Hypermetre

The responses to all three of the renditions show a pattern of response that coincides with two bar sections (a two-bar hypermeter). This is the case in bars 1, 3, 5, 7, 9, 11, and 13 but is reduced in the last bars of the excerpt. The regular accent structure (creating the hypermetre) seems to be one of the features contributing to the phrase perception. Towards the end of the excerpt, the regular accent structure is contradicted by the phase-shift of the melody. The decrease in responses and their structure in the second half of the excerpt may have its source partially in this less regular structure. The difference in the proportion of the response at each of the positions of the first half of the excerpt however, indicates that this is not the only feature that is contributing to the phrase perception. In bars 5, 7 and 9 especially, there are other features that seem to contribute to a much higher response.

A smaller proportion of responses also coincide with one-bar sections. Again, the regular accent structure (i.e. metre) is one of the features contributing to the perception of phrase structure. This lower level structure is discerned even when the explicit cues for it are no longer there. On this level, the phase-shift does not seem to confuse the listeners. This implies that the lower level structure – once set up in the first part of the piece, keeps guiding the segmentation of the music and could help the perception of larger phrases. The structure is played with (making it interesting) but, as is shown by the listeners' continued regular response, not so much as to make it confusing.

These two levels of hypermetre may be seen as constituent parts of a larger 4-bar structure – in which we see an opening and a preparation for a close (harmonically, melodically and rhythmically).

10.6.5.2.3 The number of features and the proportion of response

There seems to be a connection between the number of features at any position and the response for PS. As discussed above (section 10.6.5.2), bars 5, 7 and 9 are the most feature intensive areas whilst also being the positions with most response. However, the relation between number of features and proportion of response is not so simple, otherwise there should be a relatively high response rate in bar 6 (with its 6 features). There may be two additional contributing factors for this. Firstly, there are not only fewer features but there are also different ones that may have less of an effect. Bar 6 shares 4 features with bars 5, 7 and 9: change in motive, change in texture, pitch jumps, and bar lines. The two new features are inexact repeats and imitation in a lower part. It does not have: following voice-leading/cadential progression, exact repeats, increase in underlying rhythm and lengthening in Gould. It does follow lengthening in Kovacevich which may explain the relatively high response in the Kovacevich recording (18 %).

Secondly, the position of bar 6; these features fall between bars 5 and 7 and it may be that their importance is further weakened by the overshadowing presence of the features of bars 5 and 7.

The greatest response for bar 6 is in the first session on first listening to the Kovacevich. Even for the Kovacevich, by the final listening the response drops. In both the Gould sessions there are few responses in any one listening. The majority of these are by listeners who press a key on almost every bar line. In terms of the metrical/hypermetrical structure, bar 6 is a weak bar.

Even though there is a relatively large number of features here, these features and their combination are not strong enough to suggest that bar 6 would be a phrase start. This indicates that the combination of features: change in motive, change in texture, pitch jumps, bar lines, inexact repeats and imitation in a lower part, despite the presence of some of these in bars 5, 7 and 9, is not a combination that elicits a perception of a phrase start. It suggests that other features: following voice-leading/cadential progression, exact repeats, increase in underlying rhythm and lengthening in Gould, which are present in bars 5, 7 and 9, but not in bar 6, are more important.

Bar 4 has a similar number of features as bar 6, but a higher response. Bars 4 and 6 share all features except the inexact repeat. Bar 4, unlike bar 6, is not between two other clear phrase boundaries. This implies that the position of the feature combination in relation to other feature combinations is important and has an effect on response, but the effect is not very large.

The music-analytic discussion above suggested that a phrase boundary could be perceived in bar 13.666. There is a response in bars 13.666-14 that could be encouraged by: following cadential progression, inexact repeats, melodic background progression and lengthening in both performances, but it is relatively low for all performances. As discussed above, it is here that the hypermetrical structure breaks down and that the syncopation and upbeats come to the fore. Furthermore, it may be that the continued descent of the melody and the overlapping imitation in the bass mean that there is no break and no new phrase. This part is an example of irregular phrase structure. Some listeners identify PS and PE while others are confused (and said so in discussion), and a large proportion do not to subdivide the second phrase at all.

10.6.5.2.4 Performances, Features and EOP

There are five main areas of EOP responses common to the responses to all recordings: the bars before bars 5, 7, 9, 13 and 17. The most restricted is in response to the Gould while the most spread out is in response to the MIDI. In term of the individual bars, the earliest response is in the expectation of the end of the extract while the shortest is before bar 13. There are some EOP responses in bar 12-12.333 in Gould and a broader range of EOP responses around this position in the Kovacevich and the MIDI. As we know from the musical analysis, the explicit voice-leading and the metrical 4-bar structure would lead us to expect a PS on bar 13. This expectation is not fulfilled however, and new expectations are expressed on bar 13 in response to the Gould (and a larger area in the MIDI and Kovacevich). In relation to the musical features, we see a lengthening in the Gould on 12.333, which could lead to expectation generation.

In bars 5, 7 and 9 the EOP responses begin from approximately a bar before the phrase start and continue up to it. In contrast there are few EOP responses for, for example, bar 6. In the Kovacevich there are EOP responses on the beat before most bar lines, coinciding with the lengthening in the performance. However, this is only the case in the first session. In the second session, the proportion is lower with a greater concentration leading up to the main boundaries. In the Gould there is a large difference between the first session and second session on bar 9 with a much reduced response on the bar line, and increased earlier expectation. The tendency for earlier responses in the second session is common to both and there are clearer areas of response in the second session. For the second session of the Kovacevich, having heard the Gould performance first, there are no responses on the upbeats to bars 2, 3, and 4. However, within each session, those who press a key at these positions do not change there minds between hearings.

The main areas of EOP responses are those that coincide with the features of voice-leading, cadential progression, changes in rhythm, motive and, in some cases, lengthening.

10.6.5.2.5 Performances, features, and PE

As a result of the experimental design, the marking of PEs must precede that of the PSs. However, the results of the PEs show the clearest difference between the performances. In response to the Gould, the PEs are, on the whole, marked on the same note as the PSs and these notes are those lengthened by the performer. On the other hand, in response to the Kovacevich, the PE is marked not only on the same beat as the PS but also on the beat before; in the Kovacevich recording it is the upbeat that is lengthened. This implies that in the Kovacevich recording the musical features are exaggerated by lengthening the notes leading up to the phrase boundary, while in the Gould, the PSs (and metrical structure) are highlighted.

The effect of the performances has been mentioned in the above discussion. The results show that the responses to the different performances are not only specific to the performance but also affect the response to the other recording heard subsequently. The Gould performance is much faster than the Kovacevich and Kovacevich highlights each bar of the first half of the piece while Gould prepares fewer phrase boundaries. As a result, the responses to the Gould are of much longer phrases while those in response to the Kovacevich also show a "subphrase" level and below. In addition, the response in the second session is affected by which recording was heard first; if the Gould was heard first the phrases marked in the Kovacevich are longer and vice versa. These results show that there is a certain phrase level that is clarified and perceived in both performances and in the MIDI; there is little ambiguity as to the importance of bars 5, 7 and 9 even though their functions differ. At the same time, there are "subphrase" levels which can be identified through the musical features and which may be accentuated by a performer. When they are accentuated they may affect the listeners' phrase structure perception.

10.7 An underlying reliance on phrases: the hidden nature yet essential role of phrasing: Wagner, *Die alte Weise* from Tristan und Isolde

10.7.1 Introduction

Wagner's music is known for its huge structures, for its long, seemingly unending lines, and performances lasting for many hours. His music is known for 'evading a solid harmonic setting' (Kerman, 1956, p. 175) and for bringing traditional Western tonal harmony to the brink. It may seem, therefore, that the idea of 'the phrase' and its contributing features so far investigated in this study may no longer be applicable for Wagner's music. However, the phrase, and its related ideas, which are used throughout many of the discussions of his music, may be key to explaining how we are able to follow the music.

The logical consistency with which, from *Das Rheingold* onwards, Wagner developed formal associations, stretching across hundreds and thousands of bars, out of the recurrence and variation or transformation of short motifs or themes, is reminiscent of Brahms.[100] Wagner and Brahms shared the central problem of how to develop monumental forms from musical ideas of only a few notes but their solutions were dramatically different (Deathridge and Dahlhaus, 1984, p. 99).[101]

Wagner's opera, *Tristan und Isolde* (1859) is often discussed.[102] In particular "*Die Alte Weise*" a Shepherd's song of Act III, has been discussed by music historians, music analysts and music psychologists and contains many aspects of Wagner's music in general while being monophonic and relatively short. It even formed the basis for a Special Issue of *Musicæ Scientiæ* (1998) and several of the articles therein are referred to here.

Die Alte Weise is a Shepherd's song performed on stage on a cor anglais following the orchestral prelude to Act III. The Shepherd's piping wakes Tristan for his first

---

[100] Just as Brahms handled the sonata form development – extending it over the whole movement, so, in Wagner's music, the formal principle of 'architectonic grouping' of components yields to that of the 'web' of motivic relationships (Deathridge & Dahlhaus, 1984, p. 99).
[101] Tristan und Isolde predates the Brahms, and Brahms is known to have studied it.
[102] With concentration on the "Tristan chord"

struggle of the act and returns at the beginning of a second cycle.[103] As a Shepherd's song we expect it to be relatively 'simple'. But this is not a 'typical' folk song (Court, 1986, p. 18).

The motivic relationships, as well as the harmonic, voice-leading and rhythmic characteristics of this relatively long melody generate, confound and satisfy our expectations and thus create our sense of phrase of the piece.

10.7.2 The piece

Die alte Weise (henceforth, Wagner) consists of two contrasting ideas: one based around a duple construction forming the main theme and the other based on triplets.

The opening theme begins with a rising 5th in minims giving the tonic and dominant notes of the excerpt, establishing $\hat{5}$ as the primary tone (Forte, 1998, p. 19). The ascent continues to the e^{b1} before a descent down to a^{b1} (bar 4) the last two notes of which are again minims, providing a rhythmic 'frame' for the opening four bars. For Forte (1998), the arrival on a^{b1} completes the descending linear progression that prolongs the primary tone within the first four bars. In these bars both pulse and key have been established (two of the features we expect from the beginning of a piece or section). Nattiez (1998, p. 46) points out that the first four bars are a phrase not only because of the 'courbe de liason' and of melodic and rhythmic 'repos' on the mediant a^{b1}, but also because of the contrast with the material that follows.[104]

In bars 5 – 9 syncopations within the bars play against the pulse. In bar 8 even the articulation of the first beat of the bar is lost. Instead, the upbeat to bar 8 (i.e. the off-beat) is highlighted by an appoggiatura. The next note to fall on the beginning of the bar only arrives in bar 10. In terms of voice-leading and key, the descent

---

[103] The curtain rises, showing, according to the stage directions, a castle garden on a rocky height, with the sea visible here and there as far as the horizon. '[T]he whole place suggests the absence of a master; … in ruinous decay'. In the foreground Tristan, seemingly lifeless, lies on a couch in the shade of a great lime-tree; at his head sits Kurneval, bending over him in grief, listening intently, anxiously, to his breathing. 'From without comes the sound of a Shepherd's pipe' (Newman, 1961, p. 264). Despite the relatively simple genre of a Shepherd's song, for Newman at least, the long unaccompanied cor anglais melody is one of the strangest and most poignant ever imagined by man. 'It accelerates towards the end in a series of triplet turns and at last dies out mournfully in the lowest register of the instrument' (Newman, 1961, p. 264). For Negus, it is 'a mixture of the sophisticated and the natural' (Negus, 1993, p. 24).

[104] For Chailley 'Elle [the phrase] atteint la thésis SOL en deux groups arsiques et se résoud sur une katalèse; broderie par le LA bémol, inflexion sur "un SOL bémol inattendu souligné d'un sforzando …mais pusique le sol bémol a été identifié comme simple "note de passage", ce n'est pas lui qui est "inattendu", mais le sforzando qui le souligne!' (1972, p. 96, in Labussière, 1992, p. 46).

begun in the first four bars is taken up again with a return to b*b* in bar 5. Two attempts are needed for the a*b*^{1} – g^1 descent thus extending the length of the phrase. When the g is reached for the first time (in bar 7.25) all the performers lengthen it relative to the previous beats, thus accentuating it. In the Barenboim recording the g is lengthened, in the de Waart the previous beat is also lengthened, and for the Böhm, both the previous and the following beats are lengthened.

The descent continues to g*b* in bar 8.75 creating an expectation for f in bar 9 (which would give an even-numbered 8 bar opening). However, this arrival is delayed, by a bar-long *gb^1*, to the beginning of bar 10. Instead of a simple arrival on f, the chromatic g*b* receives special emphasis.[105] The arrival and hiatus on this g*b*^{1}, its displacement of the 'expected' f, usurping its stronger hyper-metrical position, and its relative length compared to what has come before, does shed temporary doubt on where we were aiming. Lerdahl sees the *gb^1* slightly differently: 'On the one hand, the $\flat\hat{2}$ induces a strong attraction to $\hat{1}$, to which it resolves at the *beginning* of the next phrase … On the other hand, the g*b*^{1} acts as a discordant $\flat\hat{5}$ within the implied dominant chord' (Lerdahl, 1998, pp. 32-33). Lerdahl connects the rôle of this g*b*^{1}, which extends the phrase by one bar, with later use of it in phrase extensions, culminating in bars 31-4.[106]

In bar 10, f^1 is finally reached and the opening material returns, modified in the third bar, delaying the previous arrival on a*b* until the beginning of the next bar (bar 14).[107] The a*b* of the end of bars 1 – 4 is reached, again on a minim. Looking more closely at bars 10ff there is an exact repetition of bars 1 to 3 (3rd quaver). Following this, the rhythm of bar 3 is retained but there are melodic changes. The minor second underlined in bars 8 – 9 is repeated in the 'lower' part of bar 12.

The a*b* now falls on the first beat of the bar, not half way through a bar as it did in bar 4. It is also part of a falling fifth, which begins an inversion of the opening material.[108] Two of the performers lengthen the first note of the inversion, (Barenboim and de Waart, graph 3.6.6.5, appendix 3.6.6). Although this a*b* signals the beginning of a melodic inversion (the rhythm is the same as the original), and could therefore imply the beginning of a new phrase, this does not correspond to

---

[105] 'Cette nouvelle note est cruciale puisqu'elle n'a pas encore été entendue, elle dure cinq temps et, surtout, elle crée une tension vers la tonique fa qui ne se résoudra qu'avec la dernière note du passage' (Nattiez, 1989, p. 47). Later in the act, when the music is repeated, Tristan begins the phrase '*Muss ich dich so verstheh'n…?*' with the same *gb^1*. According to Forte (1989, p. 19), this *gb^1* is associated with Tristan's persona and is a Schenkerian 'interruption'.

[106] 'in a tormented version of the *Sehnsucht* motive' (Lerdahl, 1989, p. 33).

[107] The two-part texture of the opening is heard more clearly and remains for much of the melody. The 'upper' part has the same descent from (c) – e*b* – d – c – b*b* – a*b* as bars 1 – 4 with a relatively long b*b* before a*b* is reached in bar 14. The lower part, descends from (f) – a*b* – g – g*b* – f – f*b* – e*b* – d*b*.

[108] When this section is repeated later in the act, the falling fifth sets the keyword *Sterben* (to die).

a tonal beginning. The a*b* here is part of a larger descent (or rather 'ending') through g (bar 20), g*b* (bar 36) to the final f (bar 42). The exact inversion of the opening continues to 16^1 where the dotted crotchet – quaver pattern, extended earlier, is developed and combined with triplets.

In contrast to the previous 'falling' lines, bars 15 – 20 move back up. This change happens after the c of bar 15 and it is with this note that a pedal begins in the 'lower' part.[109] The new beginning suggested by the motivic inversion of the opening material and the minim start in bar 14 was weakened by the harmonic implications. Now, in bar 15, the 'primary tone' (c, Forte, 1998) is picked up and emphasised with the pedal. The overall 'direction' is changed: not only is there a general rise, this rise includes within it a return up the a – g – g*b* line of bars 11 – 12 and in the 'upper part' of bars 15 – 16 (Nattiez, 1998, p. 49). Furthermore, a new motive results from the combination of previous ones. Bar 5 was preceded by two minims, with the new beginning following them. If a new beginning is felt anywhere in bars 14 –15, it may be in bar 15.25 rather than on 14. Nattiez, identifies the inversion of bars 10 –11 but emphasises that 'la phrase s'achève par un repos' on $\hat{5}$ (bar 15) which for him marks the PE. Not only is bar 15 the PS, Nattiez states that this would, in an 'analyse traditionelle', be the beginning of a development section involving the development of melodic material of bars 1 – 9, which constitute, for him, the first two phrases (Nattiez, 1998, p. 48). For Deliège's subjects, whose responses are discussed in chapter 8.3, the phrase boundary falls between the c and f of bar 15 and this is one of the phrase boundaries that the principles she uses can predict (Deliège, 1998). All three performers lengthen the middle two beats of bar 15.

The 'upper' line of bars 10-13 led to a*b* and there is now concentration around a*b*–g–g*b* leading to the g♮ of bar 20. The 'lower' part of bars 10-13 concurrently leads to d*b* and then C in bar 15. In bars 17-19 there is concentration around D*b*–C finally resolving on an (octave displaced) C^1 in bar 21.

For Lerdahl, however, the second phrase extends from bar 10 to 21. However, the a*b* of bar 14 has 'a double function' due to grouping overlap; it ends the line beginning on c in bar 10 and also begins the line arriving on g in bar 18. Lerdahl also highlights that this a*b* is contained within a larger prolongation from c to g (Lerdahl, 1998, p. 33). Labussière also sees bars 10-21 as the second phrase, which is, for him, like a Schenkerian prolongation of bars 1-9 (Labussière, 1992, p. 48).

Bars 22-26 see a development of bars 5 – 7. Again the 'upper' line leads down from e*b* to g, a line first seen in bars 2 – 5 (Lerdahl, 1998, p. 33) with emphasis on the c^1, while the lower part leads down from g to c (reaching it in bar 27). There is also a parallel descent at a third below the upper line. Nattiez highlights a different

[109] The dominant pedal is not unique; Wagner's music is described as having a '[l]ove of dominant pedals' (Truscott, 1963, pp. 83).

descent including the triplets from c in bar 22 to the b♮ in bar 26 leading to c in bar 27.

The dominant c and its dominant g are emphasised (bars 20, 21, 24 and 26-7, along with high e*b*). Unlike the previous pedal points, which concentrated around c and g, this time (bars 26.25-30.5) only g and c are sounded, clarifying the dominant and its own dominant. According to Nattiez, the alternation between c and g does nothing but prolong the two notes which the previous phrase had led too (Nattiez, 1998, p. 51).

In bar 30.75 the rhythm is increased with the triplets following each other more closely. Here the g*b* of bar 8.75 returns, and in addition to the g*b*, f and e♮ are introduced to the triplets emphasising the tonic with its nearest chromatic auxiliary notes – as Nattiez says, the g*b* and e♮ both play the same structural role in this key pulling towards f (and yet avoiding a resolution on it). The long g*b* (reminiscent of bar 9) returns again in bar 36 leading to the f in bar 38 via a snippet of bar 5, which Lerdahl calls a bridge (Nattiez, 1998, p. 33). Bar 38 begins as though it will be another inversion of the opening material but instead is a variant of it, which leads down to the lowest note of the extract, the low F, the last four notes of which highlight the three note descent a*b* – g – f. The last five bars can therefore be seen as the final phrase.

Another phrase division can be identified. For Nattiez, the 'development section' ends in bars 35-7 with the return of essential earlier elements: the g – c triplets rhythm in bar 35, the g*b* with its relative length, the tritone descent between the end of bar 35 and the beginning of bar 36, the return of bar 24 in bar 37 which contains a melodic descent g*b* – f – e♮. For Nattiez, then this is a two bar reprise but he includes it as part of a longer phrase. For him, bars 32 – 42 form a coda. He points out a reprise of the initial melodic pattern f – c – a*b* of bars 1 – 4 in which he seems to be referring to the f on bar 38, c in bars 39 – 40, a*b* in bar 42.25 leading to the final descent a*b* – g – f. This descent summarises the descent of a*b* (bars 4, 7 –8) – g (bar 8, prolonged by g*b* in bars 8 – 9) ending in the last bar with f (Nattiez, 1998, p. 52).

Moreover, Nattiez suggests another possible segmentation: one that would delay the beginning of the final phrase to bar 39.75 following the repose on the dominant c underlining the connection between bars 38 – 9 and bars 14 – 15.

This discussion of some of the voice-leading, rhythmic and harmonic characteristics has touched upon the phrase structure and implied phrases of the extract. Broadly the first bars of the extract have a relatively clear phrase structure setting up a framework that is developed or worked against later on in the extract. Several phrase structures have been proposed by different theorists and in an experiment. Most of the discussions lead to the identification of phrase boundaries. As will be discussed below, when combining these theoretical discussions with listeners' responses and performance contours it seems that it is not so much the phrase boundaries themselves but the expectations of them assist us in following this music.

The listeners' responses and performance contours are briefly introduced and followed by a comparison of these with the analyses above.

10.7.3 Listeners' Responses

Graphs 3.6.6.1-4, appendix 3.6.6, show that though the responses for single positions can be relatively low for the MIDI, the responses to all renditions have 1) areas of PS, PE and EOP responses which together have a large response, 2) the relationship between the responses, like for the other pieces, varies from the EOP preceding the PE, the EOP and PE coinciding and preceding the PS and all three coinciding, and 3) the positions chosen in the different renditions are, usually, the same.

10.7.4 Performance Contours

Tempo contours

Graph 3.6.6.5 in appendix 3.6.6 shows the tempo contours for the performances. It shows that the tempo changes are mainly in the same positions and in the same direction and often to the same degree. There are some exceptions such as bar 37 where Barenboim's contour shows a lengthening while the others hardly do, or bars 32-33 where the note lengths increase in the Barenboim and decrease in the other two.

Intensity contours

The intensity contours of the Barenboim and De Waart recordings are different. However, for the most part the minima are in similar bars including 4-5, 9-10, 15, 17, 22-23, 26-27, 32 and 35. The minima, however, can be of different lengths between the performances (graphs 3.6.6.6-8, appendix 3.6.6 and chapter 7).

10.7.5 Listeners' responses, performance contours, analysts' phrases and features

Here the results of the listening studies are discussed in more detail in comparison with the performance contours and the phrasing discussed by the theorists (section 10.7.2). A comparison with Deliège's results is discussed in chapter 8. The graphs are presented in appendix 3.6.6. The discussion begins with the positions most often discussed by the analysts (section 10.7.5.1) followed by an analysis of the positions with most musical features.

10.7.5.1 Phrase positions and features in view of the analysts' discussions

Bars 1 and 9-10

There are two phrase divisions that all 8 theorists agree on (bars 1 and 10). At bar 10 the opening theme returns in the same key and register as the opening and follows the PE features on bar 9. As in the Brahms, repetition of the opening seems to be one of the strongest PS indicators.

A PS could be expected on bar 9 and is by some listeners in all the renditions. For the MIDI, there are EOP responses mainly throughout bar 7 and the PE responses are spread across bars 9-10. The responses cover a similar area for the performances: the EOP responses continue clearly into bar 8, and the PE responses are on the long note of bars 8-9. In terms of tempo, graph 3.6.6.5, appendix 3.6 shows that all three performers lengthen the long g*b* relative to the beats before and after it in preparation for the new phrase. In terms of intensity, graphs 3.6.6.6-8, appendix 3.6, all have clear diminuendos just before and some in bar 10.

Bars 4-5 and 22

The next most popular positions are bars 22 and 5. Five analysts identify the first PE at the end of bar 4 and the next PS on bar 5. This could be because bar 5 is seen as phrase boundary at a lower level in the hierarchy than bars 1 or 10 (bars 1-4 and 5 – 9 forming a two-phrase pairing within the larger phrase of bars 1-9). However, those analysts that do identify bar 5 do not explicitly say this. For others, there could be no phrase boundary here. As discussed above, in bars 1 – 4 there is a descent down to $\hat{3}$. By bar 10 the descent to $\hat{1}$ is complete. Schenker's first chord label appears at the start of bar 4, implying that harmonic motion only begins from bar 4. Following Rothstein, this implies that a phrase cannot end here. The 'descent' of bars 5 – 9 is more explicit than in bars 1 – 4, and the motivic material is less 'thematic' and has a more 'closing' character. In terms of 'function' then, are bars 5 – 9 'closing'.

MIDI listeners identify PE and PS during bars 4-5 but a higher proportion of listeners identify EOPs in this area than the specific location of PE or PS. Unusually amongst the case-study pieces, the EOP location seems clearer to listeners than PE or PS.

All three of the performers increase the tempo on bar 3.25 in preparation for a ritardando which starts on bar 3.75 and continues until the upbeat to bar 5. For all the performances, this is the position with the greatest tempo change and the position with largest decrease in intensity since the opening is just before bar 5.

Bars 21-22 and 20

Six analysts identify a PE at the end of bar 21 and PS at the start of bar 22. Such a (relatively) large number here is, to some extent, surprising. The theme was inverted in bar 14, so rhythmically (if not melodically) a PS may be expected there but only Forte identifies this bar as a PS. The c^1 at the start of bar 22 is tied to the previous c so there is no clear metric stress at the start of the bar. Moreover, the voice-leading descent led to the g♮ in bar 20 and the new start, the return to the high c^1 is in bar 21, a bar earlier than the new phrase identified by the theorists and a position that is only identified by one analyst (Lerdahl, 1998).

On the other hand, bar 22 sees the return and variation of the motivic material of bars 5ff (see above). However, more analysts choose bar 22 than bar 5 and four of the six that choose bar 22 also choose bar 5 and one of the five that chooses bar 5 chooses bar 20 and not bar 22. Deliège, who identifies both bars 5 and 22, states that this can be explained for 'motivic' reasons (Deliège, 1998, p. 67).

MIDI listeners identify an EOP, PE and PS on bars 20 and 21. The PE response on bar 20 and the PS response on bar 21 are the highest number of listeners choosing any single note throughout the piece. The PS at 21 only strictly coincides with Lerdahl's analysis. Listeners also identify a PS on bar 22.

All three performers lengthen the beat before bar 20 and all three increase the tempo at bars 21 – 22. Two performers then slow down through bar 22 and lengthen the first beat of bar 23 dramatically. The de Waart recording lengthens the last beat of 22, which coincides with Deliège's phrase boundary in this bar.

Bars 15 and 17

Three analysts identify the sixth quaver of bar 15 as a new phrase. This comes part way through the inversion of the opening theme and is explained by the attainment of the c (V/F) on the previous note and is the beginning of the 'ascent'. As discussed above, Forte identifies the PS at the start of the inversion (bar 14) while (Cotard, 1895) identifies the PS on the second.

Few listeners identify a PS on the 6th quaver of bar 15 though more listeners identify a PS somewhere in the bar. Some identify an EOP on 14. This precedes the PE which has a spread of responses, peaking on 14.75 and 15.25. In Deliège's experiment on the other hand, almost all subjects identify the second beat of bar 15 as a phrase boundary (possibly due to performance cues, section 10.7.2).

In the performances, the middle two beats of bar 15 are lengthened there is then an accelerando leading to the first note of bar 17 which is also lengthened. This is a different contour to that of other phrase boundaries in the piece such as in preparation for bar 5 where there is a riterdando before the longest note and then the tempo picks up immediately, or there is a big difference in tempo before the previous note and the long last note of a phrase (such as in bar 9). In bar 15 however, it seems that the long note is enabling the accelerando to follow; the different type of structural boundary is expressed in a different contour.

Bar 14

No listeners identify a PS on the beginning of the inversion in bar 14 but some do on 14.5. Similarly, a relatively small proportion of listeners respond here for the performances. As described above, there are EOP responses here (section **Bars 15 and 17**). Bars 14 – 15 are the first bars where the analysts described above disagree about the phrase boundaries. Here the listeners show that they expect the end of the phrase but identifying where exactly is more difficult. Two of the performers lengthen the first note of bar 14 and the third keeps the tempo steady. In comparison with previous PS areas, here the tempo fluctuations of similar magnitude as those of the surrounding notes (and smaller than both bar 5 and 10).

As discussed above, there are several possibilities for the structure of bars 14 – 19. All the performers use tempo to highlight parts of bars 14, 15 17 and 20. However, each performer also uses a greater riterdando in one position than the others in these bars.

Previously, a longer note has been interpreted as the last note of a phrase preparing for a new phrase. In those cases the last note has coincided with the last note of a bar and at least a temporary tonal and voice-leading closure. In bar 17 we have reached the goal of the ascent but this is not the same type of tonal closure and the long note falls on the first beat of the bar and starts off the new 'pedal' of the next two bars. In this case the lengthening occurs on the beginning of something new as well as the goal of the previous melodic line and seems therefore to be a different kind of phrase boundary to the one identified at bars 9 - 10.

Bars 26-28 and 29-32

Throughout bars 29-32 the triplets start and stop, each time "attempting" a different variant looking for the 'right one'. Each time they generate expectations with different possibilities as identified by different analysts. Nattiez and Cotard identify the g*b* at bar 30.75 as the PS, while Deliège identifies it at the beginning of the following bar. The scheme shown in Labussière (1992) shows the first beat of 32 as the PS. In PS responses to the MIDI and performances, bars 31 and 32 also have a higher response than bar 30.

Bars 26-28 have a similar scattering of PSs. In bars 30-31 Nattiez identified the PS at the beginning of the tied g*b* while Deliège saw it as occurring one crotchet later at the beginning of bar 31. Similarly in bar 26 Nattiez saw the phrase beginning at the start of the long g in bar 26 while Deliège sees it as beginning during the same note but at the start of the next bar (bar 27). It seems that metrical position overrides other aspects for Deliège while melodic characteristics are overriding this for Nattiez. All three performers lengthen the first beat of bar 26 leading to Nattiez's PS on the second beat. As there is no change in the note on the first beat of bar 27 it is not possible to conclude from the tempo whether a PS should begin there. In bars 19 – 20, 25 – 26, and bars 34-35 the triplet figure is used in preparation for the end of the phrase and the listeners' EOP responses coincide with this.

The tempo changes here vary. For example, in de Waart's recording, tempo changes occur within the phrase seemingly to accentuate the rhythmic, melodic and harmonic features and eventually prepares and leads us to the PE.

Bars 36-39

Bars 36 – 39 also have a scattering of PSs but this time all the analysts choose at least one PS in this area and some (like Nattiez) even explicitly discuss the ambiguity. There is a clear PE peak on 36.5 with fewer responses 36. The EOP responses peak in the same position (36.5). The PE and EOP seem more pinpointed than the PS. There are differences between the responses to the performances at these positions (graph 10.7.3.2).

Bar 37 is considered by some analysts as a PS, by others as the bar after a PS and by still others as preceding the first bar of the last PS of bar 38. It seems from the EOP responses that the last possibility is identified by some listeners.

All three performers lengthen the two beats before bar 36. The tempo contour of bar 35 is very similar to that before bar 5 but this time, after the phrase boundary, the tempo is much more steady. In the Barenboim recording the beat before bar 38 is lengthened even more than that before bar 36 but in the other two recordings, although there is a lengthening, the difference is much smaller. From bar 39 to the end of the extract there is a gradual but dramatic riterdando in all performances.

There are again EOP responses in bars 39-41 (especially 41). Listeners may remember that the end of the extract is approaching. Moreover, there are indications that the end is nearing: the register is lower than it has been and the line is descending.

10.7.5.2 The Features

As with the other pieces, the above discussion elaborated upon a number of musical features. Here the most feature intensive areas (with seven or five features) are discussed the rest are given in the feature graph 3.6.6.9, appendix 3.6.6.

Seven features: Bars 9-10

The most feature intensive area is on the first beat of bar 10 (7 features): complete tonal motion, exact repeat, change in motive, long note, following voice-leading, following cadential preparation, bar line. The responses to all renditions are amongst the highest in the piece.

The new phrase starts at the start of bar 10. The phrase is prepared during bar 9: the voice-leading, implied harmony and the expectation for a symmetrical 4-bar structure. In the MIDI responses results in both positions are equally chosen indicating that the musical features have equal effect for anticipation and actual PS.

In general, there seems to be a greater coincidence between the music features and the responses to the de Waart recording than to the Barenboim. The PE during the note beginning on the up beat to 9 is clearly marked in both performances.

Five features: Bars 4-5 and 14

The first beats of bars 5 and 14 have five features. The first beat of bar 5 has change in motive, four bar template, following voice-leading, following cadential preparation and bar line. The first beat of bar 14 has following voice-leading, inexact repeat, bar line and lengthening in Barenboim and de Waart. For both of these there are more EOP responses than PE and PS which peak a bar later.

The responses in bars 4 – 5 reinforce some of the notions discussed for bars 9 – 10. The proportion of listeners responding to the performances at this position is a little lower than that of bars 9 –10. In that of MIDI there is almost no response. The marks for the PE are clearly given at the end of bar 4 for the De Waart and Barenboim recordings. As in bars 9-10, the response to De Waart fits most closely with the PS though the difference with Barenboim is smaller here.

The MIDI response is much smaller in bars 4-5 than in 9-10. The features observed on bar 10 and not on bar 5 are complete tonal motion, exact repetition, long note. There are no features that are observed in bar 5 and not in bar 10. Of the five features in bars 4 – 4.5 and 8.75 – 9, there are only two unique features: four-bar template in bars 8-9 and following cadential progression in bar 4. The difference in response between MIDI and performed recordings in bars 9-10 implied that the performance was enhancing different cues and clarifying the decision in one direction or another. This factor could also partially account for the difference between the responses to the Barenboim recording and those to the MIDI in bars 4–5. It could also be that without performance features, the other features are not felt – the overriding features of complete tonal motion and then repetition of bar 10 are not present in bars 4 –5 and without these, performance cues are needed as an aid for further subdivision.

Two features are common to both 4.75 and the first beat of bar 5 (cadential preparation and voice-leading). On 4.75 there is also lengthening in Barenboim (preparing for 5) and on 5 there is also a bar line, as well as a change in motive and four-bar template. Just like with the PS, there are two main peaks around bars 4–5 for the EOP from bars 3 to 4 in both recordings. There are many more responses to bar 3 of the de Waart recording in the first session than the second. The responses to the Barenboim recording in the second session are higher at the start of bar three and lower half way through the bar. It may be that the listeners are affected by having heard the de Waart recording first at the start of bar 3. At the end of the bar however, the trend is the same for both recordings: there are more responses on first hearing than second. This may reflect a learning (and unlearning) process. The features given at the end of bar 3 – the descent of the melodic line, the diminuendo, the nearing of the fourth bar and the implied harmony may suggest that the phrase end is nearing. However, in this Wagner extract, our expectations are played with and the listeners (by the second session) learn that, despite the features indicating the approaching PE, it will be delayed. There are similar peaks in the responses to the MIDI recording.

The main theme of the piece is inverted in bar 14. Here there are 5 features: bar line, lengthening in both performances, inexact repeats and explicit voice-leading. However, there is very little response to this. It seems that, as the theorists themselves say, the repeat (the inversion) that can be identified in theory (from the score) is not similar enough to the original to be identified and to encourage the sense of a PS. Instead, the nearest PS is identified a bar and a half later, after the descent has finished and for the beginning of the new ascent and theme. The response here is greatest for the Barenboim recording even though there is relative lengthening in both. The De Waart responses over bars 14 –15 for the EOP are equally spread and are similar for both sessions. In response to the Barenboim recording there are equal responses in bar 14 and bar 15 with the second session having higher responses than the first. The MIDI responses for the EOP are spread over bars 14 –15. In response to the De Waart recording, the PE is marked on the first beat of 14 as a peak among a spread of responses. A larger peak is on the second beat of bar 15, which is, surprisingly, larger in the first session than in the second. In response to the Barenboim recording on the other hand, there is only a peak on the second beat of 15 and the response increases the more it is heard. Further feature areas are discussed in appendix 10.7.

10.7.6 Major characteristics of phrasing features in the Wagner

There seem to be two categories of features to which there are responses in this piece: motivic and harmonic/voice-leading. There are situations in which there are mainly (or only) harmonic features (such as bar 15), which may even contradict motivic ones and vice versa. Both seem to play independent roles in encouraging phrase decisions. Responses are greater when the two coincide. There are positions where the metrical structure coincides and there are others where it is offset but does not seem to play the most important role. Some of the features primarily prepare PEs, such as the voice-leading progressions (high EOP responses) and others confirm, such as the repetition (higher PS responses).

There are some areas for which the response to the De Waart recording is large in session I and the response Barenboim is large in session II. There other areas where the response is equal (bars 7.5 and bar 9) where the musical features seem to be clear regardless of performance or listener; in bar 7 there is the change in the sequence and it is in the middle of the harmonic and melodic progression leading to the phrase end. In bar 9 the end of these progressions is finally reached.

Overall, the responses to the three renditions are usually in the same areas. However, within the positions there are some differences. There are a number of positions (including, bars 4 –5, 9 – 10, 15 – 16, 17, 22 – 23, 24 – 25, and 32) where there are at least two peaks and where the majority of the responses to the Barenboim are concentrated on the first peak while the majority of those in response to the de Waart recording are on the second peak and small responses in both positions in the MIDI. In these cases, listeners seem more prepared for PSs in the Barenboim recording than in the de Waart. There are also positions where there are responses only to the Barenboim recording such as bars 12, 35 and 39. These positions are usually the end of bars and subdivide longer phrases into subsections. They are in the areas of implied cadential progressions and sometimes of lengthenings in the recording. There are also a few areas where the trend between nearby peaks or the choice of a single position is similar such as bars 20-21,26-27, 31, and 40 – 41.

As with other pieces these results imply that there are some areas where the performer (in this case the Barenboim recording) prepares the listener a little earlier for the PS. The areas in which responses are made only to the Barenboim recording show the importance of the performance in bringing out cues for lower-level phrasing. There are other areas however, where the musical features are not affected or are affected in the same way, by the different performances resulting in similar response rates.

10.7.6.1 The performance contours in relation to the analysts' discussion

There are changes in tempo in all of the areas and positions discussed by the analysts. The largest decreases occur just before a PS and at the PS the tempo usually picks up again (such as in bars 9-10). At these positions all three performers usually have very similar tempo contours. It seems that there are more commonalities in tempo changes in accentuating phrase boundaries than in other parts of the extract.

In areas where different musical features highlight different phrase boundaries (as in bars 14ff), the tempo contours of individual performers seem to match most of the different positions to an extent but to highlight one, clarifying the one structure that dominates over the others in each performance.

Changes in also tempo occur during phrases, for example, pushing forward towards phrase end. This often further accentuates musical features (as in bars 32 -36). The listeners' MIDI responses coincided with the musical features, but in performance these are further accentuated. The performer can also accentuate particular structural features (like the arrival on bar 17 and the start of the ornamented dominant pedal), helping orientate the listener within the phrase.

This discussion has shown that tempo is used in different ways to highlight various aspects of the music but it is difficult to use tempo as a sole predictive tool to identifying phrasing. In combination with other musical features, tempo change helps to clarify PS, PE and EOP. Changes in tempo can also accentuate other musical features that may not be phrase parts but that contribute to our interpretation of the structure and therefore our interpretation of PS, PE and EOP.

10.7.6.2 Expectation: applicable, or even essential

In this discussion, the complexities of the Wagner and the different interpretations of its structure and expectations have been explored. The phrase structure of the opening of the piece, setting up the framework for the rest, is relatively clear. Later in the piece, phrase boundaries that are identified by the analysts are in the same 'areas' but not the same notes. The results from the listening experiments showed that there is a tendency to begin to expect the end of the phrase in the same areas showing that there are features which are causing expectations in many listeners. These include:

- repetition of melodic/thematic material
- descent to $\hat{1}$
- hypermetrical structure
- dominant pedal
- triplet motive and its acceleration

The expectations generated stem from features that are not new: the generation of the expectation of the tonic through harmony, rhythm and voice-leading and the frustration of those expectations through delay. There are features, such as the different use of harmony and the expanded structures that are often treated as new with Wagner. In terms of the expanded structures 'the vast increase of size in Wagner's music affects everything, down to the most incidental of harmonies, which, occupying one beat or two at the most in a normal sized work, can easily in Wagner's proportions become a matter of several bars, so that the whole of one's hearing apparatus must be adjusted to deal with this music if one is not to derive the most erroneous impressions from it'. Even the expansion is not new 'but was already existing in Schubert, a good deal of whose new methods of tonal and harmonic progress in his instrumental music are based on greatly enlarged passing phrases, usually founded on one or other of the Neapolitan harmonies; what is new with Wagner is the size to which he has expanded these processes. But they never leave the line Schubert started' (Truscott, 1963, p. 82).

Truscott explains that 'all th[e] unquestioned individuality of thought on Wagner's part is, like most of Beethoven's, so far from avoiding or suppressing such a plain foundation, dependent on the simple perfect or imperfect cadence. And indeed most of Wagner's harmonic and tonal innovations, many plainly anticipated by Schubert, consist simply of elongating notes which in an earlier composer would fall (or rise) almost immediately. Wagner delays this process until he has spun out our sense of suspense to a thin fine thread and eventually we go where we originally expected to go; or he uses a note as a hinge on which to go somewhere else and gives us the process over again – an extension of the enharmonic change coupled with the simplest idea of sequence. The long *appoggiaturas*, the Neapolitan flat supertonic and flat submediant – without these, extended to great length *dependent upon eventual conformity*, there would be no characteristic Wagnerian flavour. This, of course, does not explain his very real and very great genius, but it does expose those of his basic methods which help most to impart that flavour' (Truscott, 1963, p. 77).

In the Wagner, different voice-leading, harmonic and rhythmic structures can be identified. These different possibilities can result in an ambiguity; one person may identify one, other or several structures, possibly creating confusion. This piece can be interpreted differently depending on which features are focussed on. However, although the resulting phrase structures may vary, the expectations of phrase boundaries (covering greater area) implied by the descriptions of the analysts and explicitly expressed by the listeners seem to be common.

10.8 Summary of case-studies

One of the main aims of this study is to explore the question of musical phrasing from a number of different directions (the combined approach). In these case-studies ideas from music analysis, general music theory, music psychological literature, specific musical analyses, performance characteristics, listener responses, and written responses have been considered. This combination has enabled an investigation of where the different theoretical approaches overlap, where they cover different aspects and how these relate to participants' responses.

Music analysis and the listeners', musicians' and performers' results combined

Musicological analysis helps the identification and description of individual musical features as well as the more complex combinations. This approach enables the suggestion of a number of phrases and musical reasons for their perception. Lack of complete agreement, even between music analysts, shows that there is no one right answer.

In many cases, listeners identify the same positions as those identified in the musicological analysis. In these cases, the features identified in the musicological analysis can often explain responses. When there is co-occurrence of phrase identification between participants, performance contours, and the results of analysis, this allows the suggestion of clear (systematic) musical reasons for the responses. However, there are phrases (such as the inverted repetition of the Wagner) identified by some analysts that are not responded to by listeners. This mismatch indicates that musicological analysis alone cannot account for all and only listeners' phrase identification.

Text

Some theorists highlight the importance of text and it often seems possible to relate different aspects of the musical structure to the text structure or content. The text therefore seems important for the musical structure. However, when asked, the majority of listeners were not listening to the content of the words (for the performances, the majority could not understand them, and there were none in the MIDI versions). The text therefore may be a 'structuring' device, having its (often simultaneously different structures) mirrored in the musical structure, but the meaning of text, and the grammatical structure often do not seem to (consciously) play a rôle for listeners. This does not mean that the phonetic structure of the text is not perceived and used.

Differences between populations

The detailed case-study approach allowed the general differences and similarities between populations observed in earlier chapters to be analysed in more detail.

Perhaps unsurprisingly, the degree of similarity between populations seems to depend on the degree of clarity of the musical features: the clearer the feature combinations, the more similar the responses. Moreover, factors other than musical experience, namely the musical features, seem to be more important in eliciting different responses.

Comparison between responses to MIDI and two performances

Again, the detailed case-study approach allowed the general differences (in proportion) and similarities (in location) between responses to MIDI and performances and between performances, to be analysed in more detail. The differences can often be related to specific performance features, themselves related to the musical ones.

The combination of the performance studies and the listeners' responses suggests that there is often a relationship between the performance features studied and the listeners' responses, as has often been observed. In particular, there does seem to be a relationship between phrase-final lengthening and diminuendi and phrase ends.

The results indicated that the degree of similarity between performances seems to depend in part on the range of options given by the musical features. In cases in which the performance features of different performances differ, they sometimes emphasise different length sections (such as the differences between Gould and Kovacevich in the Brahms), and sometimes emphasise the same general areas but do so in different ways (such as lengthening different beats within the same phrase end in the Bach Suite).

Although, it was observed that there is often a riterdando and/or diminuendo towards phrase ends, there may also or only be lengthening and an intensity minimum at end of phrase and there can be note-lengthening at start of phrase (Brahms). There are also examples of clear crescendo and decrescendo patterns and sometimes accelerando and riterdando patterns within phrases. However, these are not the only patterns observed; performance features also coincide with other musical features (such as metrical, textural/pitch accents). Not all changes in tempo and intensity are directly related to phrasing (even though the patterns can 'look' very similar).

With attention to the tempo and intensity contours' characteristics, it seems that these can indicate phrase identification. They will therefore be returned to in chapter 15.

Feature analysis and experimental results combined

The combination of the different approaches allowed for the identification and analysis of musical features. Some features seem more important than others for phrase identification, they seem to differ in function, and there appear to be different combinations of features eliciting different kinds of responses.

Beyond the phrase parts that listeners were asked to identify (PS, PE, and EOP), the results indicate that other, more detailed phrase parts may be important aspects of phrasing. Many theorists have mentioned different phrase types (chapter 9). The results indicate that it is also possible to relate the different features to phrase types. These topics are discussed in more detail in chapter 11.

Chapter 11

Frequency of features, their combinations, and occurrence with phrase parts:
Expectation of the end

11.1 Introduction

The analyses of patterns of listeners' responses, score annotations, performance contours, and analysts' descriptions (chapters 2-7), and their discussion (chapter 10), showed that all of these can be directly related to musical features, and that the proportions of listeners' responses can be related to performance features.

This chapter presents an analysis of the musical features that occur at certain positions both alone and in combination, and their relationship with the phrase part responses[110] and more detailed phrase parts emerging from the analyses of the responses and musical analyses reported in previous chapters (chapters 3-7). The results of the analyses in this chapter contribute to the characterisation of the different musical features and phrase parts which is used for the discussion of musical features and phrase-parts (chapters 12 and 13) and is partly summarised in a rule base and algorithm (chapter 14). This chapter also includes a discussion of the performance features of tempo and intensity change and listeners' responses (section 11.3.6).

[110] Phrase Start (PS), Phrase End (PE) and beginning of the expectation of the phrase end (EOP).

11.1.1 Aims

To assess:
1. The relative importance of features and their combinations as phrase-part indicators.
2. The relationship between the types of features and the phrase-parts and types with which they occur.

11.1.2 Hypotheses (derived from chapter 10)

The responses analysed in this study indicate that features differ in importance both individually and in their combinations.

a) Some features (and their combinations) occur both with and without phrase-parts,
b) Some features occur more frequently with than without phrase parts.
c) Some features (and their combinations) occur only with certain phrase-parts.
d) Different features occur with different phrase-parts and phrase-types.
e) Features that occur with phrase-parts occur least frequently.

11.1.3 Method

This study consists of four parts:
a) Determination of the frequency of each feature in the case-study pieces.
b) Calculation of the proportion of occurrences of each feature with PSs and PEs out of the total number of occurrences of this feature.
c) Calculation of the proportion of occurrences of each feature in a phrase part out of the total number of phrase parts (PS and PE).
d) Observation of the locations of features in relation to areas of EOP response.

A computer program was developed for these calculations.[111]

Features were first analysed in relation to the PE/PS responses and then in relation to more detailed phrase parts. As the Wagner excerpt is almost thirty bars longer than any of the other case-study pieces, the relationships arising from the Wagner would dominate in this analysis and is therefore reported separately. However, results were found to be generally similar.

11.2 Results: Features coinciding with major phrase part responses

The determination of the frequency of each feature in the case-study pieces and calculation of the proportion of occurrences of each feature with PSs and PEs out of a) the total number of occurrences of this feature, and b) with each phrase part,

[111] Many thanks to Dan Tidhar for his help here and discussions about the rule base and algorithm of chapter 14.

leads to a categorisation of features according to their frequency and frequency of occurrence with phrase parts both individually and in pairs (sections 11.2.1 and 11.2.2). This also leads to an analysis of the minimum features necessary for the identification of the phrase parts of the case-study pieces (section 11.2.3). A more specific analysis, investigating the relationship between different phrase parts and features is also possible on this basis (section 11.2.4). In these sections, the analysis is based on phrase part areas that had high responses (major phrase part responses), summarised in section 11.2.5. In the following section areas with lower, but systematic response are analysed (section 11.2.6).

11.2.1 Individual features occurring with phrase parts

Tables 11.2.1 and 11.2.1a show the number of occurrences of each feature in the case-study pieces. Comparison of the number of occurrences of each feature with the location of phrase-part areas identified in chapters 3-10 shows that approximately half of the features occur with phrase-parts for more than 80% of their occurrences (category A) and that that those features that occur most often, occur less often with phrase parts (category B). The features are presented in their categories in tables 11.2.1 and 11.2.1a.[112]

[112] The feature numbers were assigned at an early stage of the research and are presented for ease of reference. They do not indicate relative frequency or importance.

Table 11.2.1: Three feature categories (for five pieces), showing for each feature **A) Total number of occurrences of each feature,** **B) Total occurrences of feature with phrase part,** **C) Proportion of features occurring with phrase parts (% of features), and** **D) Proportion of phrase parts occurring with features (% of phrase parts)**					
Feature No.	**Feature**	**A**	**B**	**C**	**D**
Category A: Features that occur more often with a phrase part					
2	Following cadential progression	26 (11 resolved)	23 (11 of resolved)	92 (100 of the resolved)	41
3	Exact repeat	2	2	100	0.04
6	Following explicit voice-leading	14	14	100	25
8	Four-bar template	13	12	92	21
13	Following increase in rate of harmonic rhythm	11	10	90	18
14	Four-bar template from beginning of melody	17	14	82	25
Category B: Features that occur more often outwith phrase parts					
1	Bar-line	59	16 (& 6 on half bars)	27 (37 including half bars)	28
4	Inexact repeat	17	6	35	39
5	Imitation in lower voice	7	0	-	-
7	Pitch jump	44	28	63	50
9	Long note and/or rest	55	17	31	30
10	Implicit voice-leading	13	7	53	12.5
11	Change in texture	35	16	46	28.5
12	Change in motive	14	6	42	11

Table 11.2.1a shows that for the Wagner the categories contain similar features. There are three features that do not appear (mainly due to the monophonic texture). There are two features that are in different categories: change in motive is more important and four bar template is less so.

Wagner 11.2.1a: Feature categories for Wagner alone. For each feature A) Total number of occurrences of each feature, B) Total occurrences of feature with phrase part, C) Proportion of features occurring with phrase parts (% of features), and D) Proportion of phrase parts occurring with features (% of phrase parts)					
Feature No.	**Feature**	**A**	**B**	**C**	**D**
Category A: Features that occur more often with a phrase part					
2	Following Cadential preparation	13	11	85	44
extra	Complete tonal motion	2	2	100	8
3	Exact Repeats	1	1	100	4
6	After 'explicit' voice-leading	11	9	82	36
12	Change in motive	7	5	71	20
13	Following increase in rate of harmonic rhythm	4	3	75	12
Category B: Features that occur more often outwith phrase parts					
1	Bar lines	42	9	21	36
4	Inexact repeats	12	1	8	4
7	Pitch jumps	16	2	12.5	8
8	Four bar template	11	4	36.4	16
9	Long notes	16	6	37.5	24
Features that do not occur in Wagner					
10	Melodic background progression	--	--	--	--
11	Change in texture	--	--	--	--
5	Imitation in lower voice	--	--	--	--

The following discussion begins with the most frequent (Category B) features. The less frequent, more 'reliable' (Category A) features follow in section 11.2.2.3.

11.2.1.1 Individual Category B features

Bar lines (1)

Bar lines are among the most common features, representing the strong beats in predictable metrical units. The results show that as they occur so often, they are generally not useful as predictors of phrase boundaries and sometimes coincide with other sections of the phrase such as the 'climax'.[113] They may, however, indicate the type of PS (anacrustic or down beat), or PE (strong, masculine cadence, or weak, feminine cadence), though even this is not always the case (as in the Mozart Sonata example, section 10.3).

Inexact repeats (4)

On their own inexact repeats are unreliable.[114] However, if they are preceded by another feature, these combinations occur consistently with PE/PSs (section 11.2.3.3).

Imitation in lower (accompanying) voice (5)

As exact markers of phrase parts, this feature is relatively unreliable.[115] It seems that in order to be useful, repetition should occur in a clear voice.

Pitch Jumps (7)

For Lerdahl and Jackendoff relatively large pitch intervals (not followed by another large pitch jump, thus avoiding momentary accents) indicate grouping and phrasing boundaries (Lerdahl and Jackendoff, 1987, see also chapter 8.2). Moreover, the 'relatively more pronounced' pitch intervals indicate boundaries at higher grouping levels. Although the level of the phrase and what makes an

---

[113] Sixteen phrases occur with bar-lines (and 6 occur on half-bars). Even if the bar-line was taken to indicate proximity rather than exact coincidence with a phrase-part, only about half (or two thirds if half-bars are included) would occur in such relation to identified phrase parts. Similarly, in the Wagner, 9 out of 42 bar lines occur with PE/PSs.
[114] Six inexact repeats coincide with phrase-parts out of 17 total inexact repeats and 56 total PE/PSs. Of these, 4 inexact repeats coincide with PS (out of a total of 21 PSs). In the Wagner, the situation is even 'worse' with only one out of 12 exactly coinciding with a PE/PS.
[115] Out of the 7 occurrences of imitation in the lower voice, none occur exactly with PE/PSs identified though there are several cases in which imitations in the lower voice, occur very close to PSs.

interval 'relatively more pronounced'[116] in Lerdahl and Jackendoff's theory are not clear, the implication is that the larger pitch intervals should indicate phrase boundaries. Throughout this section 'PE/PSs' have been referred to. However, pitch jumps are amongst the few features that really do relate to the phrase 'boundary', the interval being the 'gap' between the PE (the first note of the pair) and the PS (the second) (section 11.3).

The current analysis shows that there are some intervals that only occur without PE/PSs while others do also coincide with PE/PSs. However, only one interval (that of two octaves) coincides with PE/PS more often than not.

Within each piece, the following intervals occur only with phrase parts (including thirds or larger):

Table 11.2.1.1.1: Intervals occurring only with phrase parts

Piece	Interval	Occurrences with phrase parts
Bach Passion	Rising octave	3
Bach Suite	Rising 5th	1
	Rising octave	1
	Falling octave	1
Mozart Sonata	Falling 6th	1
Mozart Aria	Rising 4th	1
Brahms	Rising 6th	1
Wagner	Rising 4th	1
	Falling 5th	1

In the Bach Suite, where there is a lack of other features (such as explicit harmony),[117] one rising fifth, and one rising and one falling octave seem to be relatively important for PE/PS identification. In the Bach Passion, where there are several complicating factors (polyphonic, complex harmony) there are also three rising octaves that coincide with PE/PSs. In the rest there are two or less intervals larger than thirds coinciding with PE/PSs.

Graph 11.2.1.1.1, appendix 11 shows the percent occurrence of each interval (ascending and descending) and the percent occurrence of the intervals out of the total number of phrase parts identified in the listening studies. It shows that most of the intervals that occur in these pieces are represented at phrase parts. However, most of these intervals do not occur exclusively at phrase parts. Unlike in the descriptions of the Gestalt principles (chapter 1) and the implication of

[116] GPR 4 (Lerdahl and Jackendoff, 1987, p. 346). This rule could express, for example, intervals that are more distant in actual pitch, in terms of key or in terms of interval type. For example, minor thirds are experienced as closer than major sevenths, while octaves are experienced closer than both.
[117] In a manner similar, to some extent, to the situation in the folk songs (for example, Bod, 2001 and Ferrand et al, 2003, discussed in chapter 8).

Lerdahl and Jackendoff's GPRs, it seems that the larger intervals do not coincide only or even more often with PE/PSs. This applies both for size and type of interval. The smaller intervals occur more with the phrase parts (except, not surprisingly, the rising fourth). Moreover, the proportion of PE/PSs which occur with the same note repeated is the second highest. In other words, cases opposite to a pitch jump are very common markers of PE/PSs (see also Bod, 2001, and chapter 8.5).

From the discussion of other studies (chapter 8) and these results it seems that, though pitch jumps sometimes coincide with identified PE/PSs, they occur more often elsewhere. Furthermore, it seems difficult to clearly identify pitch intervals (and their direction) that are reliably indicative of PE/PSs, especially as the same interval does and does not coincide with PE/PSs. For example, the falling fourth in bar 2 of the Mozart Sonata coincides with a PE/PS while the falling fourth in bar 9 of the Bach Passion does not.

This analysis indicates that using pitch jumps to predict PE/PS positions is problematic.

Long note/rest (9)

Unlike pitch jumps, a larger proportion of long notes coincide with phrase parts as is clearly seen in graphic form (graphs 11.2.1.2.1.2, appendix 11). Usually the long note is identified or treated as the last note in the phrase. However, as the graphs show, sometimes there is a higher co-occurrence on long notes with the penultimate note in the phrase (i.e. lag 1 note). Table 11.2.1.1.2 gives correlation values (R) testing both cases.

Table 11.2.1.1.2: Correlations between long notes and responses

	Response on same beat as long note (R)	Response note after long note (R) (i.e. lag 1 note)
Bach Passion	**0.675**	0.262
Bach Suite	**0.760**	0.332
Mozart Sonata	0.098	0.002
Mozart Aria	**0.873**	0.144
Brahms	0.611	**0.820**
Wagner	**0.6738**	0.161

Graphs 11.2.1.2.1.2 and the correlations (table 11.2.1.1.2) indicate that there are many situations in which relatively long notes coincide with PE/PSs either during the relatively long note or rest or following it. The correlation is closest in pieces that have fewer other clear musical features, such as the Bach Suite, the Mozart Aria or the Wagner. In the Wagner, the harmonic structure is complex all the way through and so relatively long notes are important. In the Mozart Aria and the Bach Suite the harmony is relatively static, so the long notes again become important for boundary detection. In the Mozart Sonata the longest note does

coincide with the clearest phrase part. However, the other phrase parts coincide with 1) a note whose relative length is common in this piece or 2) a note whose length is the second shortest of the piece. For the latter, this is preceded by two relatively long notes and these may be contributing to the preparation for the PE (and therefore EOP responses). In the Bach Suite the longest notes occur with the first note of the phrase rather then at the boundary between the end of the previous phrase and the new one. However, like in the third phrase boundary in the Mozart Sonata, two of the boundaries are also preceded by notes that are longer than the notes that precede them, again maybe helping prepare for the phrase boundary. There is also high correlation in the Brahms, in which the 'long notes' return systematically on every bar line (chapter 10).

These results and the discussions in chapters 8 and 10 indicate that it may be useful to distinguish between long notes and rests, and broaden the view to see relatively long notes as indicating either the preparation for a phrase end (such as Bach Suite bar 3), the phrase boundary (such as Mozart Sonata, bar 4) or the start of a new phrase (such as Bach Suite bar 2). Conversely, however, there are several cases for which long notes or rests do not occur with phrase parts.

In general, relatively long notes and rests may not be regarded as clear PE/PS indicators. However, if there is no other information (for example, if the harmony is static or too complex) long notes and rests become critical features (section 11.2.3.3).

Melodic background progression/Implicit voice-leading (10)

Seven occurrences of the 13 melodic background progression coincide with phrase parts. As with the inexact repeats however, if these coincide with other features, the proportion of occurrence with PE/PSs rises (section 11.2.3.3).

Change in texture (11)

Sixteen of the 35 occurrences of the change in texture occur with phrase parts. This feature is not always relevant (such as the Wagner). Even in combination with other features there does not seem to be an increase in occurrence with PE/PSs.

Change in motive (12)

Change in motive is particularly hard to characterise. Moreover, changes in motive can occur at a PS as a 'change', but can also contribute to the preparation of a PE. In the Wagner example, change in motive is in several cases important at the start of a phrase, and for this piece, change in motive belongs to Category A.[118]

[118] Six out the 14 occurrences of the change in motive occur with phrase parts.

11.2.1.2 Pairs of Category B features

Graph 11.2.1.2, appendix 11 shows that pairs of Category B features also coincide with PE/PS less often than without, suggesting that these features can be disregarded in any attempt to identify phrase boundaries. However, there are some exceptions; in some cases the features occur more evenly with and without PE/PSs. Moreover, some combinations occur more often with PE/PSs. Such information may be useful in characterising the features indicating phrase-parts (and types) in a particular genre.

11.2.1.3 Individual Category A features

Category A features occur most often with PE/PSs. The higher the proportion of occurrence with PE/PSs, the more important/reliable the feature (table 11.2.1.1). Within Category A, exact repeats occur only twice in the case-study pieces and both times coincide with a phrase start. Similarly, 'following voice-leading' occurs only with phrase parts. However, as summarised in table 11.2.1.3, the rest of the features do not uniquely occur with phrase parts.

Table 11.2.1.3: frequency and % occurrence with phrase-parts of Category A features

Feature	Number of occurrences of features	Proportion of features occurring with phrase parts (% of features)
3. Exact repeat	2	100
6. Following explicit voice-leading	14	100
8. Four-bar template	13	92
2. Following cadential Progression	26	92
13. Following increase in rate of harmonic rhythm	11	91
15.Four-bar template (from start of melody)	17	82
10.Following implicit voice- leading	13	62

In the Wagner there is one exact repeat (a repetition of the opening). It coincides with a clear PS response. The change in motive feature, unlike in the other pieces, seems to occur more frequently with phrases than not (compare with table 11.2.1).

Table 11.2.1.3a: Frequency and % occurrence with phrase-parts of Category A, Wagner		
Feature	**Number of occurrences of features**	**Proportion of features occurring with phrase parts (% of features)**
3. Exact Repeats	1	100
Complete tonal motion (extra)	2	100
2. Following Cadential preparation	13	92 (100 for resolved)
6. After 'explicit' voice-leading	11	82
13. Following increase in rate of harmonic rhythm	4	75
12. Change in motive	7	71

11.2.2 Pairs of features with phrase parts

The results of the previous section indicate that an exact repeat or explicit voice-leading, regardless of context, always occur with phrase-parts (table 11.2.2). The rest of the features do not always coincide with PE/PSs. Instead, as table 11.2.2 shows, pairs of Category A features coincide with phrases 100% of the time.

Table 11.2.2: Pairs of Category A (strong) features and coincidence with phrase- parts		
Feature Pair	**Number of co-occurrences of two features**	**Proportion of feature-pairs occurring with phrase parts (% of features)**
2&6		
Following cadential progression, After explicit voice-leading	11	100
2&8		
Following cadential progression, Four bar template	6	100
2&10		
Following cadential progression, Following implicit voice-leading	8	100
2&13		
Following cadential progression, Following increase in rate of harmonic rhythm	10	100

Feature Pair cont.	**Number of co-occurrences of two features cont.**	**Proportion of feature-pairs occurring with phrase parts (% of features) cont.**
2 & 15		
Following cadential progression, Four bar template (from start of melody)	5	100
6&8		
After explicit voice-leading, Four bar template	3	100
6&10		
After explicit voice-leading, Following Implicit voice-leading	2	100
6&13		
After explicit voice-leading, Following increase in rate of harmonic rhythm	7	100
6&15		
After explicit voice-leading, Four bar template (from start of melody)	1	100
8&10		
Four bar template, Following Implicit voice-leading	3	100
8&13		
Four bar template, Following increase in rate of harmonic rhythm	3	100
10&15		
Following Implicit voice-leading, Four bar template from start	3	100
13&15		
Following increase in rate of harmonic rhythm Four bar template (from start of melody)	2	100

Moreover, these combinations account for almost half of the phrases and do not occur outwith phrases.

11.2.3 Features accounting for phrase parts

Here the minimum number of features needed to account for the phrase parts identified in the case-study pieces are discussed. The above results indicate that the infrequent features are most reliable. Therefore, the least frequent features are employed first.

11.2.3.1 Individual Category A features

Exact repeat (3)

As mentioned above, the 'exact repeat' is one of the few features that always occurs with phrase parts (PSs).[119]

Following explicit voice-leading (6)

As with the 'exact repeat', the resolution of explicit voice-leading always occurs with a PE/PS. Unlike the exact repeat, it occurs more often (14 times). This and the exact repeats are the most reliable features identified.

Resolved cadential progression (2)

Eleven of the 26 cadential progressions resolve, all of which coincide with PE/PSs. Of these eleven, seven are not accounted for by exact repeats or explicit voice-leading.

Table 11.2.3.1: Occurrence of resolved and unresolved cadential progressions and occurrence of each with phrase parts

Cadential Progression	Number of occurrences of features	Number of features occurring with phrase parts	Features occurring with phrase parts (% of features)
Resolved	11	11	100
Unresolved	15	13	87

As table 11.2.3.1 shows, most of the 'unresolved' cadential progressions also occur with PE/PSs. However, some of these are not 'clear' PE/PSs (section 11.3.1).

11.2.3.2 Pairs of Category A features

The above three features, regardless of the features with which they co-occur, or the more general context, always occur with PE/PSs. Similarly, if the pairs of features in table 11.2.2 co-occur, regardless of the other features with which they co-occur, always occur with PE/PSs.

[119] This feature only occurs twice in the case-study pieces. The relatively few occurrences of this feature are not surprising, especially considering the lengths of the excerpts studied.

Of the three occurrences of the combination of ‘four bar template’ and ‘four bar template from start of melody’ that coincide with PE/PSs, 2 occur with new phrases. A small but important number of PE/PSs therefore seem to coincide with the location of an ‘expected’ PE/PS. This is particularly strong when the melody starts on the first note of the piece.

Of the eight occurrences of ‘following cadential progression (unresolved)’, and ‘following implicit voice-leading’, one new phrase is accounted for.

The total number of phrases accounted for by exact repeat, following explicit voice-leading and the combinations including four bar template, and four bar template from the start of the melody and following unresolved cadential progression and implicit voice-leading is 26 out of the 56 PE/PSs identified in this study. Most of the other feature combinations of the strong pairs also occur with these 26 phrases (table 11.2.2) but do not identify additional phrase parts.

11.2.3.3 Returning to Category B

Rests (9)

Although the ‘long-notes and rest’ feature has been consigned so far to the ‘weak’ group (category B), this observation is primarily relevant for purely instrumental music. In vocal music (or where there is a clear melody distinguishable from an accompaniment, such as in the Bach Passion and Mozart Aria), the rests can be very strong features. This may be connected to their relation with breath. In the Mozart Aria, the rests allow PE/PS identification (and therefore continued ‘comprehension’ despite the lack (static nature) of harmonic and other information. In the Bach Passion, the harmonic moves are much more complicated, with the accompaniment and voice rarely coinciding and the harmony hardly resolving. Nonetheless, the clear rests in the vocal part, allow the listener to follow each ‘phrase’ i.e. the longest rest comes at the point of clearest harmonic resolution so far. There are 19 rests occurring with PE/PSs, in the two pieces (Bach Passion and Mozart Aria).

Cadential progression and long notes (2 and 9)

The remaining PE/PSs unaccounted for are PSs. These follow resolved cadences and usually coincide with inexact repeats (Bach Suite and Mozart Sonata). They were excluded from the above discussion because the PSs do not directly follow the cadential progression but are temporally separated by a long note. Again, under these conditions, the long notes (and sometimes pitch jump information) may be useful. However, the rôle of the long note here is prolongational.

11.2.4 Phrase Parts

The different feature categories and feature combinations (section 11.2) can be related to individual phrase parts explicitly identified by participants and through analysis of performance contours (PS, PE, EOP), and more detailed ones emerging from these analyses (such as prolongation of the end).

11.2.4.1 Phrase End

The analyses and results of this study indicate that there are different parts of phrase-ends, including arrival, and prolongation of the phrase end (which may be followed by another arrival), are typically preceded by the preparation for the end of the phrase (the beginning of the expectation of the end of the phrase).

For all of these parts, the largest single group of PEs is accounted for by cadences; of the 25 cadences, 22 coincide with PEs. The three that do not coincide are unresolved. However, there is not a simple division between resolved cadences (that do coincide) and unresolved (that do not) as there are, in total, 15 unresolved cadences in the case-study pieces. Not only do these, in a small proportion of cases, not occur with phrases, but also cadences do not account for all PEs. However, combinations of cadential progression with all the other strong features do seem useful:

Table 11.2.4.1: Feature pair combinations and the number of times these pairs coincide with identified PEs

Feature Pair	**No. of occurrences of feature pair**	**No. of occurrences with phrase end**
2&6		
Following cadential progression, After explicit voice-leading	**14**	**13**
2&13		
Following cadential progression, Following increase rate of harmonic rhythm	**10**	**10 of which 3 are new**
2&10		
Following cadential progression, Following implicit voice-leading	**8**	**8 of which 4 are new**
2&8		
Following cadential progression, Four bar template	**6**	**6 of which 1 is new**
2&15		
Following cadential progression, 4-bar template (from melody start)	5	4 of which 0 are new

2&Rest		
Following cadential progression, Rest	2	1 of which 0 are new
2&3		
Following cadential progression, Exact repeat	1	1 of which 0 are new
(8&15)		
4-bar template, 4-bar template (from melody start)	8	5 of which 1 is new

In the one combination of "Following cadential progression, After explicit voice-leading" that does not coincide with a phrase end, the voice-leading is in the accompaniment and if the accompaniment is played on its own (or primarily listened to) a PE could be expected. With this in mind, the first four combinations in the table above (Table 11.3.2.1) can be taken as reliable and useful combinations; they only occur with PEs and they each add information about phrase parts which is not given by the combinations listed above them.

The next three combinations in table 11.3.2.1 (Following cadential progression, Four bar template (from start of melody), Following cadential progression, Rest, Following cadential progression, Exact repeat) may be supportive but not indicative. These may be regarded as additional to the other features, and not exclusive PE markers. The last entry in table 11.3.2.1 indicates that the four-bar template is most useful when the melody begins at the start of the piece.

The feature combinations listed in table 11.3.2.1 account for 22 of the 34 PEs. Ten further PEs coincide with rests, mainly in the vocal pieces. Four of these follow other features that occur a beat or two earlier. One coincides with a resolved cadence (and a bar-line) and is the only time that a resolved cadence occurs without any of the strong features of table 11.3.2.1. One PE occurs with the end of the Bach Suite extract. In the Wagner, PEs coincide with resolved or unresolved voice-leading and implied cadential progressions, pitch jumps, and the ends of sequences.

11.2.4.2 Beginning of the expectation of the end (EOP) & Beginning of the end

It seems that the expectation for phrase endings plays a rôle in phrase identification. For example, there are several responses at positions before which expectations could have been generated, but which were thwarted (such as in the cases of unresolved cadences). There are different contributing features to the expectation of the end, which differ according to the temporal span they affect and therefore the type of expectation they seem to contribute to. The comparison of EOP responses (as well as the PE responses) and the feature graphs (appendix 3.6) indicated that:

In some cases predictive features and listeners' EOP responses begin long before the PE. The listeners expect the PE well in advance, relatively long term expectations:

1) The expectation may begin at gradually over several notes. Evidence is gathered over several notes and responses for the EOP are spread. These usually involve cadential and voice-leading progressions (Brahms bar 5) or repeated motive patterns that become associated with the end of the phrase (Bach Passion, bars 7-8 and 12, 3.5, 8.75-9.0). Even if the actual location of the end is not clearly predictable, there is a general expectation. These EOP positions had large responses in the listening studies, extending over several notes.

2) Though the expectation may last several notes, in some cases the position of EOP is clear and the PE is expected early in the progression by many listeners. Here there are standard cadential and voice-leading progressions, including I-IV-V-I and $\hat{3}$ –$\hat{2}$ – $\hat{1}$ (Mozart Sonata, bars 3, 5, 7). The preparation can be relatively short and unelaborated (including IV - V – I, Mozart Aria, bars 7 and 15.5).

3) 'Preparation' can also be much shorter (last minute) relying on one chord (Brahms b. 9) or short melodic preparation motives (descending third semiquavers in Bach Suite, upbeat to bars 2 and 3) and are often combined with stronger confirmation features (such as repetition in Brahms bar 9 and Bach Suite bar 3, or 'gap' features such as the long notes of bars 2 and 3 in the Bach Suite).

These can be combined. The long-term preparation is reinforced by a last minute one. For example, the long-term preparation of bars 8-9 of the Mozart Sonata is reinforced by a change in texture and harmonic rhythm and additional ornamentation.

11.2.4.3 Prolongation of the end (and the end of the end)

Arrival at the resolution (section 11.3.2.1) often signals a PE. However, there are several cases in which a PE is reached (usually on a strong beat, with the end of a cadential progression) and then prolonged, either by more notes within the same beat or by extension into the next (usually weak) beat (just as Koch, 1787, 1983 suggests). In these cases, the (mainly predictive) PE features often occur by the initial arrival at the end. Instantaneous features after the prolongation end confirm and reinforce the PE. There are 7 such cases in table 11.3.2.1. These include: 1) four cases of rests (Bach Passion), 2) two inexact repeats, and 3) end of piece (Bach Suite). In table 11.3.2.1, these different parts of the same PEs are counted individually (so seven PEs appear twice in the table). These characteristics seem to account for some of the 'areas' of response.

11.2.4.4 Phrase Start and Gap Phrase

There are 34 PSs. The main features that occur with PSs are: repetition in the melody, change in texture, and four-bar templates (see table 11.2.4.4) of which repetition is the strongest. In other cases, it seems that the PS is most dependent on the previous PE. Indeed once the EOP has begun, a new PS is usually expected. The clearest PE-PS is when such a beginning is confirmed by a repetition.

Table 11.2.4.4 showing for each feature:
A) Number of occurrences of the feature,
B) Its occurrences with PSs,
C) Occurrences of the combination of the two as percent of total number of each feature,
D) Occurrences of the combination of the two as percent of total number of PSs (of 34).

Feature	**A**	**B**	**C**	**D**
Exact repeat	2	2	100	5.8
Unresolved explicit voice-leading	8	4	50	11.7
4-bar template	17	8	47	23.5
4-bar template from melody start	13	6	46	17.6
Rest	24	11	45.8	32.3
Voice-leading	14	6	42.8	17.6
Cadential progression	26	10	38	29
Resolved cadential progression	11	4	36	11.7
Increase in rate of harmonic rhythm	11	4	36	11.7
Change in motive	14	5	35.7	14.7
Resolved explicit voice-leading	6	2	33	5.8
Inexact repeat	17	4	23.5	11.7
Change in texture	25	7	28	20.5
Chord	35	8	28	23.5
Pitch Jump	44	11	25	32.3
Resolved implicit voice-leading	5	1	20	2.9
Position of start	61	10	16	29
Bar line	59	9	15	26.4
Melodic background progression	13	2	15	5.8
Long notes	28	4	14	11.7
Unresolved implicit voice-leading	8	1	12.5	2.9
Unresolved cadential progression	15	6	4	17.6

PSs can follow a number of different phrase parts: arrival on a clear end, following a prolongation or a 'boundary', or even an 'unclear' PE. Broadly, they can be grouped in two categories: There may be a gap (brought about by the 'gap' features) between the previous PE and the PS (here referred to as 'delayed') (14 out of 34 PSs). The PS can follow the PE immediately or coincide with it (here referred to as 'immediate').

The PSs in the Wagner are dominated by change in motive. This occurs with 7 PSs and for an additional PS the motive is modified rather than changed completely. There is one exact repeat (of the opening) which is locally a change in motive with respect to the preceding bars. Five new PSs begin with long notes 5 (of which 2 are not accounted for by change in motive). One PS occurs only with a pitch jump.

The above discussion indicates that the different features occur with, and indicate different phrase parts, more or less clearly.

11.2.5 Summary of Features coinciding with major phrase part responses

In general, the results of this analysis show that features can be categorised according to their frequency of occurrence with phrase parts: Those that occur more frequently (or only) with phrase parts (Category A above) and those that occur more frequently away from (or equally with and away from) phrase parts (Category B).

Within Category A, some features occur only with phrase parts, and are infrequent overall in the case-study pieces. While their rarity seems to contribute to their success as phrase-part indicators, their occurrence alone cannot always be relied on.

The features within Category B (particularly long notes and rests but also implicit voice-leading and others) cannot be disregarded completely. There are situations in which these features do more reliably signal phrase parts through: combinations with other features (such as cadential progressions), lack of other (Category A) features in the area or piece in general, or because of a specific genre (such as the importance of rests in vocal music).

However, difficulties remain with these features. For example, rests having the same length can, in the same piece occur both with and outwith phrase parts. Excluding rests, Category B features alone may not clearly signal phrase parts, but they are usually present with Category A features and have enough of an association with phrase parts to 'confuse' listeners (section 11.3.5).

Single and pairs of strong features (Category A) occur with 26 of the 56 PE/PSs identified in this study. A further 19 occur with rests in the vocal pieces, and the rest of the phrase parts are the starts and ends of extracts.

This categorisation is based on frequency of occurrence in general and occurrence with phrase parts. However, another categorisation may also be useful; one that describes them in terms of where they occur temporally, when they 'impact' most, and what their 'function' is in relation to phrase parts (section 11.3.1). Furthermore, the relation between features and phrasing has been described here in terms of phrase 'parts' in general. However, the results show that the features are more specifically related to the different phrase parts, which are now discussed in more detail.

11.2.6 Relatively minor response positions

The previous sections of this chapter have concentrated on PE/PSs recognised by a relatively large proportion of listeners, musicians, analysts and sometimes performers. Listeners' and musicians' responses for other areas are fewer, less consistent and vary between listenings (see also chapter 4). Here examples of these are discussed showing their importance in identifying how features change in importance according to context (they are all described in more detail in chapter 10).

In the Bach Passion there are many examples of relatively few but close responses (appendix 3.6). These seem to be related to features in the vocal and orchestral parts that do not coincide. For example, in bar 11 the vocal line has a PE on the end of beat 3. However, in the accompaniment the 'next' PS is already on beat 2. During the MIDI rendition in particular, listeners respond to one or the other in a given listening and so overall the response is lower but systematic.

In bars 4 and 6 of the Bach Suite inexact repeats coincide with lower responses. There are few preparatory features here and there are other (stronger) combinations of features nearby. It seems that even though there are stronger features nearby, some listeners still systematically respond to these features.

In the Mozart Sonata fewer listeners identify the third phrase PE/PS than the others. Features that would encourage a response here include the interrupted cadence, the symmetry with position of the phrase identified by most listeners in bar 2. However, the last note of the 'phrase' is much shorter than the last notes of the previous phrases (lasting only a semi-quaver) and is immediately followed by more semi-quavers. The closing features and predictive 'template' feature are contradicted by the immediate continuation (or lack of confirmatory features).

In bars 1-4 and 9-12 of the Mozart Aria where there are falling octaves followed by rests (overlapping between melody and accompaniment), there are relatively fewer but systematic responses which indicate the strength of such closing features. For some listeners (during some listenings) it does not seem important that the resulting sections are very short, consisting mainly of closing features.

In the Brahms, there seem to be competing features with lower response in some areas and higher responses in others. There are clear responses in bars 3 and 11 where inexact repetition is combined with others. However, the relatively low response of bar 6 indicates that inexact repeats have to be combined with, or at least not contradicted (even elsewhere) by others.

In bars 15 and 39 of the Wagner there is an end of a descent and a change in motive. Unlike for other similar positions in the piece, this signals a PE/PS for only some listeners. There is greater response for the De Waart performance in which the boundary is emphasised.

Summary of relatively minor positions

There are several different types of feature combinations that coincide with lower responses. All result in positions being identified during some but not all listenings and that may not be identified in analysis.

These features include those that:
1) When they occur (on their own or in combination), are strong enough to encourage a response while listening by some listeners but, in retrospect, do not seem to signal PE/PSs. For example, inexact repetition, changes in texture, chords (these can be referred to as online features, see also chapter 14).
2) Coincide with large responses but also coincide with smaller responses when, for example, they are temporally close to even stronger feature-combinations. For example, increase in rate of harmonic rhythm, pitch jump, long notes (these can be referred to as complementary features).
3) Coincide with large responses, but when they are not with certain other features are less strong. For example, unresolved cadential progression, melodic descent, temporal symmetry with previous PS (also referred to as complementary features).
4) When there are several features occurring in different voices at different times, usually strong features (referred to as primary features) can be weakened (especially in the MIDI rendition).

The resulting positions may be seen as internal punctuations for the phrase. As some of the complete phrases are longer than 3-7 seconds, it may be that these internal punctuations help memory, providing internal articulation of the phrase (chapters 1, 12 and 13).

11.3 Discussion: Feature Types, Phrase Parts and Phrase Types

The above results discussed the features and phrase parts primarily in terms of frequency of features, pairs of features, responses and their occurrence with phrase parts. Another perspective is to analyse the occurrence of features from the perspective of groups of features. In this section, groups of features, in the form of feature types are discussed (section 11.3.1). On the basis of this and the analysis above it is also possible to describe phrase types (section 11.3.2). Following this performance features are brought in to the discussion, and the differences between responses to performances observed in previous chapters are brought together (section 11.3.4). This is followed by a summary of the chapter (section 11.3.4).

11.3.1 Feature Types

On the basis of analysis of the case-study pieces and the responses to them, it seems that a three-facetted categorisation of the features is possible:
1. the 'temporal scope of presence' the feature covers,
2. the 'temporal scope of impact', and
3. the 'function or result' in respect the phrase part, its identification or expectation.

1. The temporal scope of presence

Different features occur over different time-spans and different parts of the features (e.g. the beginning or the end) are important for the phrase-parts. Features can be: short and *instantaneous*; from one note to the next (such as pitch jumps), or longer, *lasting over an area* with either clear/important beginning (such as exact repeats), or with the emphasis towards the end (such as cadential progressions).[120]

2. The temporal scope of impact

Not only can the 'presence' of the features differ, the temporal characteristics of their identification by the listener may also differ.
a) The short, *instantaneous* features seem to be identified almost immediately or instantaneously.
Longer features can be identified:
b) Primarily *retrospectively*. For example, only after part of an exact or inexact repeat is heard, does the listener know that there has been a repeat and whether it was exact or not. However, it seems that listeners can be very quick to make this decision whilst hearing the repetition.
c) Primarily *predictivley*. For example, once cadential progressions have begun, there can be a prediction that they will continue and end.

In most cases, the identification of the longer features seems to be both predictive and retrospective.

3. The 'function or result'

Often, features seem to be associated with particular 'functions' in the phrase: the short instantaneous features are often '*gap*' features (e.g. rests). The longer, *retrospective* features often signal a PS (e.g. the repetitions or changes). The *predictive* ones create an expectation for the PE (e.g. cadential and voice-leading

[120] This may lead to the difficulties in analysis. Here, the features are presented at the location at which they could begin to signal the new PS, such as, the second note of a pitch jump, the note after the long note, the last note of a cadence etc., (feature graphs in appendix 3.6).

progressions). However, these are not consistent for all features. For example, change in harmonic rhythm and in texture can signal PSs but also EOPs. From the results discussed above it seems that the following combinations coincide with different phrase responses (reflected in graphs 11.3.1, appendix 11).

1) When the three feature types features are present, response is often large. These seem to be the clearest, most prepared and confirmed phrase boundaries and have a spread of responses with more musically experienced listeners indicating the PS 'early' or on the first note of the PS and less musically experienced listeners indicating it on the first note or a little later.

2) When there are only *predictive* features, there is a smaller response. Some listeners, however, lifted the key signalling the PE and seemingly automatically pressed the key almost immediately. If there are predictive features encouraging the EOP and PE, listeners lift the key before PE confirmation. Some listeners explained that they became aware of this tendency but found themselves doing the same in subsequent listenings.

3) In some cases, where the responses level is between that of (1) and (2), there are often only *instantaneous* or, especially, *retrospective* features. This may be because *predictive* features usually occur over several beats and some listeners wait for confirmation before they indicate a PE. However, *instantaneous* and *retrospective* features are more confirmatory and immediate for listeners, they 'know' that no more confirmation will arrive, so it is not worth waiting for. Situations (2) and (3) are often more ambiguous.

The most common feature type is the instantaneous one and it is also this that is most commonly associated positions that do not have any phrase responses. Of the 132 occurrences of instantaneous features, 27% occur with phrase responses. Less common are the retrospective features (53 occurrences), but for this too only a minority of occurrences (34%) coincide with phrases. However, of the 26 occurrences of predictive features 70% coincide with phrases. Alone then, the predictive category features seem to be most reliable phrase predictors among the three feature-types at least in identifying whether there is a phrase or not. Furthermore, predictive features coincide in 54% of the cases with PEs, the highest coincidence with any phrase part of the three features types.

Of the three feature categories, therefore, the predictive one seems to be the most reliable, followed by the retrospective and then instantaneous. These results suggest that an instantaneous feature on its own does not predict a phrase part well enough. In some cases, the predictability of phrase parts on the basis of a certain feature is piece dependent. It remains to be seen whether these observations are generalisable (for example, to the level of genre).

11.3.2. Phrase Types

The different types of feature arrangements may be seen as leading to identification of different phrase types. The phrase types described here are concluded from the features and their occurrence with phrase identifications, but reflect the character of the phrase parts in the wider context.

11.3.2.1 End-heavy, expectation dominated, goal directed, phrase types

When there are clear closing features (cadential progression, voice-leading, increase in rate of harmonic rhythm), a PE is expected, and may be supported by retrospective features (such as PS confirmation features including repetition, change in texture and motive). However, strong predictive features do not need to be supported by such features; a resolved cadential progression and voice-leading may indicate a PE clearly enough and may dominate the phrase. Such phrases may be seen as end heavy.

11.3.2.2 Start-heavy, statement or confirmation dominated, phrase types

When there are exact repeats (or inexact repeats, changes in texture or motive and at least one of rests, long notes, position in the bar being the same as the opening), clear preceding phrase endings, are not necessary. In phrases dominated by, for example, exact repetitions, the important part of the phrase is the start and such phrases may be seen as start-heavy.

11.3.2.3 Instantaneous division, breaths or gaps, phrase boundaries

There are also cases where there are neither predictive nor retrospective features (the information is either too complex (Bach Passion) or too simple (Mozart aria). Instead, a temporal gap is the defining feature for a division between phrases. Here there is least preparation and confirmation, simply a break and a re-start; a rest or long note may be enough to clearly indicate a PE.

11.3.2.4 Combinations

Features can be combined and/or omitted contributing to more ‘complex’ phrase types including:

a) Delayed PEs; prepared but unresolved or delayed resolution, which may follow later (such as Brahms bars 6-9),
b) Equally important adjacent phrases; for example, the preparation of the delayed resolution is repeated, this time leading to a resolution (such as in the Mozart Sonata and Aria),
c) Sequence (such as inexact repeats in Wagner),
d) Elided or overlapping phrases; for elided phrases the PE and PS occur on the same note (such as some interpretations of Brahms, bars 5 and 9), while for

overlapping phrases the PS occurs before the previous phrase has completely ended (such as other interpretations of Brahms, bars 5 and 9),

e) Ambiguous phrase boundaries; refer to complex instances of overlapping phrases. The idea of ambiguity is also often used in cases where there are several different phrasing possibilities suggested by the musical features.

These are discussed in more detail in chapter 13.

11.4 Coincidence of performance features, musical features and phrase responses

The above discussion has concentrated on the musical features in relation to phrase responses. In this section, the performance features are added.

11.4.1 Tempo Change

Comparison of the tempo contours with PE identifications in the rest of this study shows that: 1) for some performances, the listeners' phrase responses coincide with tempo changes, 2) tempo changes coincide with positions that listeners responded to in the MIDI and in the score annotation tasks, and 3) tempo changes coincide with positions identified by theorists or following music analytical approaches. This indicates a relationship both between tempo change and listeners' phrase identification and between the positions accentuated in performance by tempo change and the positions identified by listeners to MIDI and by theorists (chapters 7 and 10).

11.4.2 Intensity Change

Similarly, change in intensity immediately following high intensity areas sometimes coincides with the EOP identification by listeners and the end of the diminuendo is often in the area of PEs identified (chapter 10). The responses to the performances indicated that larger and sometimes longer-lasting, gradual reductions in intensity in the performances coincide with greater listener PE/PS and EOP responses.

11.4.3 Comparison between responses to different performances – Footprints in listeners' responses

In chapter 3, section R.6.3 it was observed that there were differences between responses to different performances of the same case-study pieces. In chapter 7 it was observed that the performance contours could be very similar or different but that the difference became less significant when a performance of the piece had already been heard in a previous session. The pieces can be divided into two groups according to proportions of similarity of listeners' responses:

1. Responses to the different performances of the Bach Passion and Mozart Sonata are similar (statistically the same) in both sessions (see chapter 3, R.6.3). Though there is a significant difference between the tempo contours for both pieces, the tempo and intensity contours of the performances are still relatively similar (chapter 7.3.2.1).

2. For the different performances of the Bach Suite, Brahms and Wagner there is a statistically significant difference in the proportion of responses by the two groups of listeners in the first session at various positions (see chapter 3, R.6.3) and the tempo and intensity contours of the two performances are different (see chapter 7). The responses coincide with the contours (see chapters 3-5). Moreover, in the second session, in which the listeners have already heard one recording and now hear the other, the responses to each piece have 'footprints' of the previous responses, and the differences between the responses to the different performances are much smaller and not statistically significant (see chapter 3, R.6.3).

The analysis in chapter 4 showed that for some of the pieces there are relatively few groups of responses. Some groups usually identify a relatively small number of positions for the first listening to one recording. At the same time, responses to first listening to the other recording show far more groups each with more positions identified throughout. The identification of these groups with the method of chapter 4 indicates that there is a statistically significant level of agreement within the groups (i.e. they are not all mistakes). In the second session, the responses to the recording with the fewer groups in the first session has more positions and vice-versa for the other recording. This is even the case in the Mozart Sonata (and less so for the Bach Passion), which, overall, statistically, does not show a significant difference in responses. The most popular positions are present in all responses and are also present in the MIDI responses (see chapter 4).

The performance contours indicated that the same musical features are highlighted in all performances of each piece. The results indicate that the main phrase positions identified by listeners are identified in response to both performances and the MIDI rendition (chapters 7 and 10). In some cases, especially when the musical features suggest more than one option, the performance contours differ between performers in the extent of tempo and intensity change. Furthermore, the proportions of response to these and other positions in some cases change depending on the strength of the performance features. Moreover, having heard one performance, the listeners' responses to the second performance in the second session indicate that the performance features of the first are remembered and affect the responses to the later performance. The performance features therefore seem to become 'part' of the piece for at least some listeners.

11.5 Summary

11.5.1 Features

The analysis of the occurrence of features with phrase parts has indicated that there are several different factors influencing the importance of musical features in phrasing. The most immediately observed factor is the relative frequency of the occurrence of features with phrase parts out of the total number of occurrences of that feature. In general, the more often the feature occurs, the less it occurs exclusively or mainly with a phrase part. The more 'useful' features, occurring less frequently but usually with phrase parts (category A) include exact repeats and voice-leading progressions. Features such as bar lines (i.e. metrically strong positions) that occur very frequently, even if they sometimes coincide with phrase parts, cannot be relied on as strong indicators. Some features occur with phrase parts more often than they do not, but not exclusively. However, when these same features do co-occur in certain feature pairs, the proportion of occurrences with phrase parts rises.

A relatively small number of features occurring individually or in pairs were found to account for all phrase parts in the case-study pieces regardless of which other features were present. These features are: Exact repeat, After Explicit Voice-leading, Following Cadential Progression, After increase in rate of harmonic rhythm, Four Bar Template, Four-bar template (from start of melody) (and Following implicit voice- leading)

However, it is not possible to disregard the more frequent features for two reasons: Firstly, in some genres (such as vocal music and the importance of rests) or musical circumstances (such as when there are no (or few) Category A features in the area or piece, such as the lack of clear cadential resolution in the Wagner and the rise in importance of motivic change), these features are more important than in others. Secondly, even if they do not help with the initial identification of a phrase, these features do seem to affect the 'character' of the phrase. For example, the location of the phrase start and end in relation to the bar line has an effect on the phrase.

Another factor according to which features can be categorised is the three-faceted characterisation according to the:

1) scope of their presence,
2) scope of their impact,
3) their 'function' or 'result'.

In some cases the same feature can be in more than one of these categories but in general, features are the scope of presence is their location and duration. The scope of impact and function are: predictive (encouraging expectation), instantaneous (encouraging boundary identification), or retrospective (confirmatory). In terms of presence, the first and last usually last longer than the middle category.

11.5.2 Phrase Parts and Phrase Types

The phrase parts (PS, PE and EOP) identified by the listeners are characterised by certain features and their combination. Together with the wider musical context and their temporal relationships, the features and feature-combinations provide a comprehensive characterisation of the phrases and their relationships with the adjacent ones. Phrases can be viewed as having individual contextual characteristics because of the relative strength of features and their combinations.

EOP identification by listeners often occurs with predictive features (such as cadential and voice-leading progressions, changes in underlying tempo). Phrases that contain these features are described as having an emphasis on the 'beginning of the end' of the phrase.

PEs are usually marked by the listeners at the end (or resolution) of the predictive features. However, the strength of response varies depending on how these predictive features end and whether or not they are immediately followed by gap and/or confirmatory features. If the predictive features are resolved and followed by gap or confirmatory features, they are here described as resolved (end heavy) phrase ends. If they are not resolved and if the gap or confirmatory features occur only later, they are described as being of the delayed or thwarted phrase end type.

Phrase boundaries can be marked by gap features (such as long notes, rests or pitch jumps and are related to breath). These may sometimes occur with little preparation (few or no clear, resolved preparatory features) and may even not have clear confirmation (few or no clear, retrospective, confirmatory features). However, for the most part, these are preceded or followed by the other feature (and therefore phrase) types. Phrase boundaries, in this study, are therefore indicated by a temporal separation between the phrase end and the phrase start (such as in the Mozart Aria).

This type of relation between PE and PS contrasts with the 'elided' phrase type; in which the phrase end and start occur on the same note, such as in the Brahms. In such cases, phrase boundary features, such as long notes, rests or pitch jumps, are not necessary for the expectation and identification of PEs or the identification of PSs.

PSs identified by listeners often occur with retrospective 'confirmatory' features, such as exact repeat or change in motive or texture. Phrases that begin in this way and are dominated by these features are described as being of the front-heavy type.

Other PSs follow resolved predictive features (such as cadences and voice-leading) and gap features (such as rests or pitch jumps). In these cases, PS features are not essential for a PS to be identified.

11.5.3 Implications of relatively minor positions

Many areas in the case-study pieces have relatively small responses, particularly for the MIDI. This discussion has shown that these positions include features that are strong enough to elicit responses from some listeners, some of the time, but because they are not confirmed or clear, not all listeners identify them. These areas may (through temporary, unfulfilled expectations, or through small end, gap or start features) help in following these pieces, easing the load on short-term memory.

11.5.4 Comparison between Wagner and the other case-study pieces

The results of the analysis of the Wagner show the same trends as for the other pieces though the definition of cadential progression is less clear than in the other pieces. In particular, whenever there is an implication of a resolution (through semitone motion, or intervals of fifths or fourths to a local or global tonic) it is delayed or thwarted by later information. Expectations are generated and sometimes confirmed.

11.5.5 Coincidence of musical and performance features and phrase responses

The responses to performances indicate that the performance features seem to be responded to by listeners. This is clarified by the differences in response to different performances (with differences in performance features) of the same piece (see also chapters 7.4.3 and 10).

In general, the phrases highlighted in the performances and those identified in response to MIDI, the performances, by the analysts and by musicians on the scores are similar (see also chapter 10). Furthermore, in many cases, in the first session the performance features highlight or clarify the phrase structure for the listener. In the second session, when a different interpretation is heard (with possibly different 'clarification'), there is still an effect of the performance features of the first session.

The results of this study indicate that tempo and dynamic change (and, in some cases, breath) are important for highlighting phrases. Moreover, there seems to be a range of uses of the performance features. The comparison of responses to MIDI, the score and performances provides further evidence for Friberg and Battel's claim that performance features are primarily clarifying aspects that could also have been recognised without these cues; redundancy of this kind is a common method of increasing the efficiency of communication of information (Friberg and Battel, 2002, p. 212, see also chapters 1 and 7.2.1).

In following chapters, the different features, feature categories, phrase parts and types are discussed in more detail (chapters 12 and 13), their relationships are summarised in a rule base and re-examined with new test-pieces (chapters 14 and 15).

Chapter 12

General discussion of musical features:
The beginning of the end

12.1 Introduction

Using the 'combined' approach, the previous two chapters investigated a number of musical features that may contribute to the identification of phrases and their internal structure and, in more general terms, to our following of the music. Each feature can be seen as an example of a larger group (or category), each of which has a different rôle in phrase identification. There seem to be a number of musical circumstances under which different features and feature groups become important. These include: the characteristics of piece or genre in question, the feature combinations in a given area in the piece, musical experience or, if heard, the manner of performance. These more general categories may be applicable to western classical music, possibly non-western music and perhaps, if taken at a more comprehensive level to other media such as scene analysis of visual or linguistic information, art or text.

This study has brought together theories and approaches from musicological, music-analytic, psychological and computational literature and lead to a concentration on the following features, which were studied in more detail in the experimental sections (chapters 2-11). The features studied here are: pitch jumps, long notes/rests, change in motive, change in texture, exact repetition, inexact repetition (such as modification of the beginning, sequence, imitation in lower voice, repetition of end-motives), harmonic progressions (primarily cadences), voice-leading (both explicit and implicit), change in rate of harmonic rhythm, tempo change (referring to overall tempo and long-term tempo change and rubato), dynamics (concentrating mainly on gradual, long term dynamic change), lengths of phrase and positions of bar, and hyperbar, lines.

As discussed in chapter 11, the combined approach leads to a categorisation and understanding of the individual features, which will be discussed in more general terms here. The features are grouped into the categories according to three characteristics reflecting aspects of the features and their effect on phrase structure.

The first two characteristics are temporal: 'the temporal span of presence', a characteristic that is "identifiable in the score", describes how long the feature is present. Features can occur instantaneously (from one note to the next), or over an area (with emphasis either on their beginning or on their gradual development).

'The scope of impact' describes the scope of influence of the feature. The latter is dependent to some extent on the former but is a perceptive characteristic. Again, features again can be perceived instantaneously or over an area (with emphasis either on their beginning or on their gradual development). Each category therefore has a different kind of impact on the phrase perception. For the same feature, the temporal span of presence and the scope of impact may be different and it is for this reason that the two characteristics are used. For both characteristics, most of the features belong to only one category but some features can belong to more.

While the first two characteristics are 'temporal', the third characteristic describes the general 'function or result' of the feature. These include Gap, Change and Memory feature categories. These, in turn, prompt phrase part start, end, boundary and expectation identifications (chapter 12).

In summary, there are three aspects to each feature each with three possible attributes:

1. **The temporal scope of presence**
    a. Instantaneous, one note to the next
    b. Area, clear beginning, continues
    c. Area, emphasis towards its end (gradual)
2. **The temporal scope of impact**
    a. Instantaneous
    b. Retrospective
    c. Combined predictive and retrospective
3. **The 'function or result'**
    a. Instantaneous, gap features; boundary
    b. Retrospective, changes or repetition features; phrase start
    c. Predictive and retrospective, expectation generating features (and their resolution); expectation (and phrase end)

The following discussion of the features is organised according to these categories, elaborating on the individual feature characteristics and the changing relative importance that results from different combinations (section 11.2). The features may be seen as examples of the broader feature categories which may differ, depending on for example, the type of music. This is followed by a discussion of each of the feature categories and their 'function' in the phrase (section 11.3).

First, a summary of the features and their three category labels are presented (table 12.1). The first label relates to the 'temporal scope of presence', the second to the 'temporal scope of impact' and the third to the 'function or result'. These contain 'musical' features (identifiable in the score) and 'performance features' (only identifiable in performance).

Table 12.1: Summary of features in their categories

Label	*Scope and function/result*	Features
Musical Features:		
Presence	*Instantaneous*	Pitch Jumps Temporal Gaps: long notes and rests
Impact	*Instantaneous*	
Result	*Gaps*	
Presence	*Instantaneous*	Change in Motive Change in Texture
Impact	*Retrospective*	
Result	*Changes*	
Presence	*Instantaneous*	*Repetition* Exact Inexact including: - Modification of the beginning - Sequence - Imitation in lower voice - Repetition of End-Motives
Impact	*Retrospective* *Memory-related*	
Result	*Phrase start*	
Presence	*Gradual*	*Long term expectation generating features:* - Harmonic Progressions **- Cycles of fifths, cycles of thirds:** - Cadences - Voice-leading – explicit and implicit - Change in rate of harmonic rhythm *Short term expectation generating features:* - Delaying suspensions, - Long notes - Short rests
Impact	*Predictive*	
Result	*Expectation Generating*	
Presence	**Instantaneous**	*Standard/Ideal Expected phrase length, memory-related, 'Template' measure, through:* - Number of notes - Clock time - Number of bars - Structural or Motivic characteristics
Impact	*Instantaneous/* *Predictive*	
Result	*Template,* *Boundary or relation to template boundary*	

Performance Features:

Presence	*Instantaneous/ Gradual*	*Performance features:* Tempo change: - overall tempo, - long-term tempo change (and rubato) - short-term tempo change Dynamics - gradual or long-term dynamic change - short-term tempo change
Impact	*Predictive*	
Result	*Expectation Generating*	

Presence	*Instantaneous*	Breath
Impact	*Instantaneous*	
Result	*Gap*	

12.2 The individual features

12.2.1 Gaps

Presence	*Instantaneous*	Pitch Jumps Temporal Gaps: Long notes/rests
Impact	*Instantaneous*	
Result	*Gaps*	

Gaps in music are often seen as providing clear instantaneous information about the division between one section and the next. Most of the authors that have discussed gap features do so in relation to Gestalt principles (chapters 1 and 8). This may be because it is one of the Gestalt principles that has the clearest musical instantiations. Almost all the studies discussed in chapter 8 concentrate on these features and almost all refer to Lerdahl and Jackendoff's GPRs (1987) which rely heavily on gap features, primarily pitch intervals that are relatively large (referred to Pitch Jumps) and long temporal gaps.

12.2.1.2 Pitch Jumps

Lerdahl and Jackendoff's rule for pitch jumps assesses the relative distance in pitch between two notes within a four note context thus avoiding jumps that are only momentary and therefore are more like accents (which can be important within a phrase) than grouping boundaries (Lerdahl and Jackendoff 1987, pp. 46 and 346, see also chapter 8.1). In this and other approaches (see chapter 8), the melodic voice is used (the music studied is usually monophonic anyway). If this were done in the current study, this would mean that certain relatively large pitch intervals (such as bars 4 – 6 of the Brahms) would not be defined as pitch jumps. In this analysis, these are included. Moreover, in these approaches, all notes are assumed to be equally important. This study has shown that this may be problematic in several ways. For example, even a melody line may be perceived as consisting of more than one voice (such as in the Bach Suite and Brahms).

In the exploration of the rôle of this feature in phrasing, the same definition to that given by Lerdahl and Jackendoff was taken. From the discussion of the studies in chapter 8 and the results of the case-studies in chapter 10, it seems that, though pitch jumps sometimes coincide with phrase boundaries identified by listeners or other theorists, pitch jumps occur more often than phrase boundaries so there are many pitch jumps that do not coincide with identified phrase boundaries. Furthermore, it seems difficult to make a clear distinction between types of pitch jumps that are indicative of phrase-boundaries and those that are not, especially as often the same interval can, in some contexts, coincide with phrase boundaries, and in others does not. For example, the falling fourth in bar 2 of the Mozart Sonata does coincide with a phrase boundary in the MIDI listening results while the falling fourth in bar 9 of the Bach Passion does not.

Of all the possible intervals, eight occur with boundaries identified in the listening study (excluding boundaries that may be described as elided or overlapping). Of these, two are both as ascending and descending intervals, four as ascending intervals, one as a descending interval and one repetition of the same note (chapters 10 and 11).

It seems that, unlike in the original descriptions of the Gestalt principles (Wertheimer 1924), the 'larger' gaps do not necessarily imply phrase boundaries. This applies both literally in terms of size of interval, and harmonically in terms of type of interval (for example, minor thirds are, if taken in isolation, seen as closer than major sevenths, while octaves are seen as closer than both). The responses to the case-study pieces show that the smaller intervals are more represented at the boundaries (except the rising fourth) (chapter 11). This is observation is from a small corpus here but indicates that the relation between the size of interval and strength of 'gap' is not direct. This feature is discussed together with the other gap features of long-notes and rests below.

12.2.1.3 Long Notes and Rests (Long IOIs or OOIs)

Lerdahl and Jackendoff refer to note length and rests in three GPRs (Proximity rules 2a (Slur/Rest) and 2b (Attack-point), and Change rule 3d (length), Lerdahl and Jackendoff 1987, p. 345-6, see also chapter 8.1). These use note-length as a feature that represents change and differences in length for the distinction between groups. However, in the examples analysed in this study, it does not seem to often succeed. For example, these rules 'succeed' in Wagner, bars 19-20, Bach Passion, bars 3 and 12, Mozart Sonata, bar 6 and Mozart Aria, bars 6-7 and 14-15 (chapters 10 and 11). However, none of these coincide with phrase boundaries identified in the studies. For this rule to be more effective in this type of corpus, it would have to encompass a section longer than four notes and be more flexible.

Unlike in Lerdahl and Jackendoff's rules, long-notes and rests (IOIs and OOIs) in other studies are sometimes treated as the same, taking the inter-onset interval (the time between note onsets) as the deciding characteristic (chapter 8). The reason given is that in some cases, especially in performance, there is little difference between the performance of a long note and a long note followed by a rest, especially where relatively short rests or long notes occur.

However, distinguishing between rests and long-notes has advantages:
1) They are not always inter-changeable; it is not always possible to replace a long note with a rest and keep the same musical structure.
2) It seems that there is a difference in function between long notes and rests. For example, long notes may close a phrase while rests may indicate phrase boundaries. One of the repercussions of this view is that long notes, while not only closing a phrase, create the short-term expectation for the end, while rests often form part of the instantaneous boundary. Therefore, long notes may belong not only to the category of Presence: Instantaneous, Impact: Instantaneous, Result: Gaps but also Presence: Instantaneous, Impact: Instantaneous/predictive, Result: Expectation generating
3) Often phrases end with a long note followed by a rest at least in one part in a polyphonic piece. The phrase can, in theory, end for example at the end of the long note or at the end of the rest depending on the piece, the performer or the listener. Each of these results in important differences in the interpretations of the phrase – particularly in terms of its rhythmic structure. The lack of distinction between long notes and rests therefore obscures the different possibilities.

In this study, a note is considered long if it is longer than the notes that precede and follow (following GPR2 b). Long notes can also occur just before the end of a phrase (such as in some suspensions) or over several notes at the end of a phrase. Therefore, the results of this study suggest that long notes may indicate 1) a phrase boundary, 2) a phrase end, or 3) generate the expectation of a phrase end (section 12.2.4).

Rests of any length may indicate phrase boundaries: the longer the rest, the stronger the indication. Even if the rest does not indicate a phrase boundary, it does seem to have an effect on the perception of the phrase. The phrase may be felt as being subdivided at the rest, or, as with long notes, the rest may be delaying the continuation of the phrase.

12.2.1.4 Gap features: Summary

Pitch jumps, long notes and rests are often cited as among the most important musical features in phrase identification. This may be because much of the music studied has been in relatively short extracts and monophonic where there is little room for other features, such as repetition and change, that need more time to develop and to be perceived (discussed below). In addition, as the examples are usually monophonic, features that are more explicit in polyphonic music, such as harmonic progressions are excluded. It is possible that, with the relatively bare texture of monophonic music, these surface features are more prominent than they are in music with more voices. These may be some of the reasons why this study reaches somewhat different conclusions about pitch jumps, long notes and rests. However, even in the monophonic pieces, some of these rules have been shown to be problematic (chapter 10.3, Bach Suite).

In many pieces there are several pitch jumps, long notes and rests, but only some of them coincide with phrase boundaries. In general, these, at their most inclusive definition (i.e. all sizes), are not always reliable signals of phrase boundaries. However, there are situations where they do become important and useful signals. For example, when there is no other clear information, especially a lack of clear harmonic motion (discussed below), then a relatively long rest signals a boundary even where it may not be expected. Once the rest is heard, it may be taken as a phrase boundary, especially if supported by a later confirmatory feature (such as a change, discussed below).

This study indicates, therefore, that pitch jumps, long notes and rests are not completely reliable signals for phrase boundaries; they are among the most common musical features throughout the pieces studied and the occasions where they coincide with phrase boundaries are out-numbered by those in which they do not coincide. Moreover, when they are signals for phrase boundaries, they are often just that – instantaneous signals; the end has been reached and the new phrase is about to start. Even if they were reliable signals for phrase boundaries, these features do not explain any internal structuring of the phrase; they do not, for example, contribute to the function of generating expectation (providing 'direction') within and between phrases. Exceptions are those occurring just before the phrase end, delaying the resolution and increasing tension (see below, expectation generating features).

12.2.2 Changes

Presence	*Instantaneous*	Change in Motive Change in Texture
Impact	*Retrospective*	
Result	*Changes*	

Another feature described in the Gestalt principles is that of change or difference (see chapter 1). In music a variety of elements can 'change', including key, motive, texture, and timbre. The 'change' referred to in music studies that use the Gestalt principles is often instantaneous (such as a change in texture or register) (see chapter 1). However, others also have a predictive aspect; key change, for example, often includes a modulation. Therefore, a clear boundary on the basis of key change alone is not always instantaneous. Other features may occur instantaneously but reaction to them may be somewhat retrospective. Nonetheless, some changes (such as of texture and motive) can occur more suddenly and these are the types of changes that are taken as examples of this feature category.

Change in texture is identified by a sudden change in the number of parts or instruments playing. Change in motive is identified by a change in thematic material. It seems that the important aspect of the 'change in motive' is the change in its rhythm or pitch direction. This may be accompanied by, for example, change in rate of harmonic rhythm (see chapter 10.6, Brahms), or change in rhythm at a more 'surface' level, where the underlying rhythm remains constant, but the melodic rhythm changes (see also Brahms Intermezzo Op. 116 no. 2 in chapter 15). The rhythmic change can be in both directions: faster (as in Op. 119 no. 1) and slower (as in Op. 116 no. 2), the faster leading onwards to the new start, and the slower closing off its own.

There is often co-occurrence of changes in motive and texture. However, as changes in motive without changes in texture have been identified in this study, the two features are kept separate.

The results of this study indicate that changes in texture and motive occur with phrase boundaries almost as often as they do not. Indeed, they sometimes occur towards the end of a phrase (such as in the Brahms), highlighting the approach to a phrase end. Like the 'gap' features they are, therefore, not completely reliable signals of phrase boundaries on their own. However, when they occur with harmonic features (discussed below) they almost always reinforce a phrase boundary (there is only one occasion among the case-study pieces where they do not). They signal a change, a new beginning, and confirm the previous phrase end and boundary.

Motivic information is not only informative when it changes. Repetition of motives can also be indicative of phrase parts (section 12.2.3).

Changes: Summary

If these 'change' features coincide with or follow other (predictive) features (such as a harmonic progression, discussed below), they may act as retrospective, confirmatory features marking the instantaneous moments of the phrase end and start. However, if preceding a phrase boundary, they can also be predictive features. They cannot, therefore, be relied upon on their own for the identification of phrase boundaries. Instead they support other features and may clarify where exactly the phrase boundary falls. This double function might be confusing and may explain some of the decisions identified in this study (such as for the Brahms, bar 4, chapter 10.6).

12.2.3 Repetition

Presence	*Instantaneous*	*Repetition* Exact Inexact including: - Modification of the beginning - Sequence - Imitation in lower voice - Repetition of End-Motives
Impact	*Retrospective* *Memory-related*	
Result	*Phrase start*	

Our ability to remember previously heard music has an effect on the connections we are able to make between sections of the music. The recognition of previously heard information, especially the return to that heard at the beginning of the piece and the main subject (which are usually, but not always, the same thing) seems to be an important aspect of listening to and performing music. The length (discussed below), internal structure of the phrases and the structure of the piece as a whole affect the amount of information remembered:

> 'The importance of parallelism[121] in musical structure cannot be overestimated. The more parallelism one can detect, the more internally coherent an analysis becomes, and the less independent information must be processed and retained in hearing or remembering a piece.'
>
> (Lerdahl and Jackendoff 1987, p. 52)

There are several types of repetition that may help with long-term connections, phrasing and structuring in general.

12.2.3.1 Exact and Inexact Repetition

Exact repetition, especially of opening material, seems to be one of the strongest signals for a phrase start, even if the previous phrase has not clearly ended. The definition of exact repetition is intuitive but potentially difficult to apply in practice. In this study, the repetition has to occur at the original pitch and the beginning of the repetition has to be exact and (in most cases) last at least two bars of the original (such as in the Brahms).

The problem increases for inexact repetitions. These include sequences and repetitions where there are modifications from the beginning, retracing steps towards the end of a phrase, a closing motive and imitation in other voices. The difficulty of defining similarity and difference has been often discussed, by for example Steedman (1977), in connection with perception of rhythm and metre, and Cambouropoulos (2001), but a systematic solution is yet to be found.

[121] Parallelism is an equivalent term to repetition.

12.2.3.2 Modifications and variations on the same underlying structure

Inexact repeats can include embellishments of a theme from its beginning (such as in the Mozart Sonata). The harmonic structure remains the same but the melody is embellished. In this case, the repetition has a similar effect to exact repetition (though possibly with less immediacy or clarity): the phrase start can be identified in retrospect. In some cases the phrase start area can be identified but an exact start note is more difficult to identify if modifications have been made.

12.2.3.3 Sequence

Repetitions in a sequence (usually short sections that are individually distinct) may be identified as 'phrases' in themselves if they become identifiable units with the internal structure of a phrase. There are three aspects of the sequence that may help in its characterisation as a whole:

1) <u>The beginning</u> - The beginning may be clear (a new section) or may emerge out of the previous phrase. It seems that if the beginning of the sequence is clear, it is more likely to be perceived as a phrase start.
2) <u>The length of the sequence as a whole</u> - Whether or not this repetition is felt like the end of a clear phrase (the whole sequence) or not, may depend on the length of the sequence (in terms of beat length or number of repetitions). The shorter and/or the more structured a sequence, the more likely it would be described as a phrase.
3) <u>The last repetition of the sequence</u> - The last repetition may be a clear change and close in which case there is a clear phrase boundary, or it may dissolve into a new phrase obscuring the boundary.

12.2.3.4 Imitation in lower voice

In polyphonic music, repetition of a theme in another part (exact or inexact imitation) can indicate a phrase start in that part. This, however, does not necessarily indicate a clear phrase start applicable for the whole texture (such as in the Brahms).

12.2.3.5 Repetition of End-Motives

In many cases, the end of phrase (as little as the last bar or couple of beats) can be repeated at the ends of phrases that begin differently. For the first few times they are usually presented alongside other clear phrase-end features. Later, they become markers of phrase ends even when other features are less clear. There are two types of such cases: 1) 'new' motives are developed within piece and can be learned during listening to the piece as being associated with the end of the phrase, and 2) motives (often cadential) that are associated with phrases ends in other pieces and that are learned through listening to pieces from the same genre.

All of these repetition-types would probably be included in Lerdahl and Jackendoff's Parallelism rule (GPR 6, Lerdahl and Jackendoff 1987, p. 51, chapter 8.1).

12.2.3.6 Summary: Repetition

In this study, the position of the repetition within the phrase is taken as having functional importance and repercussions. For example, a clear, exact repetition of a previous phrase start is treated as strongly suggesting a phrase start, while a clear repeat of information that came at the end of a phrase before is an indication of a phrase end. The problem of the definition of repetition remains but in this study, as explained above, some steps have been taken to identify types of repetition and their position and functions in the phrase.

12.2.4 Expectation Generating Features

All the above features are primarily related to the phrase boundaries, starts and end. The 'stimulus' is related to an immediate 'response'. Another category of features, and one that seems to me to be the one that makes the music studied here so interesting and complex, is the category of expectation generating features.[122] These rely, for this kind of music, on western tonal harmony and, for the listener and performer to have some familiarity with it.

The temporal effect of expectation generating features can vary from long-term connections (several bars) to short term ones (a couple of notes).

Presence	*Gradual*	*Long term, expectation generating features:* - Harmonic Progressions - Cycles of fifths, cycles of thirds: - Cadences - Voice-leading – explicit and implicit - Change in rate of harmonic rhythm *Short term, expectation generating features:* - Delaying suspensions, - Long notes - Short rests
Impact	*Predictive*	
Result	*Expectation Generating*	

[122] see also, Meyer (1956); Expectation, Rothstein (1989); phrasing and complete tonal motion, and Narmour (1990, 1992); implication-realisation.

12.2.4.1 Long term, expectation generating features

12.2.4.1.1 Harmonic Progressions

There are several types of harmonic progressions often leading from one position (or area) to another. If music is heard as a motion towards a goal (and a close, however temporary), this motion occurs in stages, to different extents throughout the piece, and relates to units of different size. All 'aim' for the 'final-close' combination, but do not all achieve it. The different progressions can include long term, 'journey'-emphasising progressions: cycles of fifths, cycles of thirds, and long or short term 'end'-emphasising progressions: cadences. As discussed in chapters 1 and 9, there are several descriptions, suggested rules and theories concerning the structure of harmonic progressions and their perception. In some cases (especially within the 'music-analytic' literature), they are the basis, and almost synonymous with, the definition of a phrase. This is not surprising as it is often possible to identify phrases in the styles of music studied here using only harmonic progressions. However, there are only a few such approaches within the music-psychological, computational studies. There therefore seems to be, to some extent, a separation of features among the different disciplines. The results of the current study (chapters 9 and 10) indicate that both harmonic and other features are used.

Cycles of fifths, cycles of thirds

In these cases a journey has begun with the cycle but the exact end-point, both temporally and harmonically, may not be clear. These cycles are often the basis of sequences.[123]

Cadences

Cadences, especially those resolving harmonically and melodically onto the tonic are, in most western music-theoretic writings, closely associated with phrase endings, the definition of the phrase and can even be, to some extent, synonymous with the phrase (chapter 1, section 1.2.10.5).

There are many different types of cadences and each type includes a number of different possible chord combinations. The most straightforward and usually strongest type of cadence is the 'perfect' cadence (see chapter 1.2.10.5). During the dominant and pre-dominant chords (or the substitutes) expectation for the end of the phrase begins to be generated. From here, there is no turning back but there are several possibilities. If the I is reached (harmonically and melodically), the cadence is fully resolved and the end of the phrase is usually also reached. However, the arrival on the tonic may not occur at the same time in all parts, or

---

[123] These are generally more common than in the particular case-study pieces analysed here.

may be extended in all parts after it has first been sounded (Bach Passion). The position of the next phrase start is not necessarily automatically predictable from the cadential progression. A phrase may start in the area from the dominant chord to either when the harmony changes, or when there is a clear feature for a new phrase start, or when there are clear features of phrase end such as I and $\hat{1}$.

If the cadence does not resolve as described above it can have the same chords but not resolve to the tonic in the melody (or less often the bass-line). The final I chord may be replaced with the relative minor (VI) (an interrupted cadence), or the progression may end on the V chord (an imperfect cadence).

For such cases, the certainty of the phrase end is reduced but the 'function' is not obliterated. If combined with phrase start or phrase boundary features, a new phrase may still begin either on the chord following dominant or on the final chord of the progression or the one following. The character of the phrase however, would be different, usually necessitating a completion (consequent) or repetition with a different ending.

Alternatively, if no such start or boundary features follow, the phrase, although having had the expectation for a phrase end, would not be complete in which case neither a phrase end nor a phrase start would be heard, and would, instead, be replaced by a continuation of the previous phrase. This, however, does not mean that these 'incomplete' progressions have no importance. These features generate our expectation for the end of a phrase, and the mere expectation has an effect on the continuation. These features enable the phrases to be extended by introducing expectation and delay (generating interest), and hiatus 'mid-way' or towards the end of a phrase.

As with the other features, the effect of cadences (especially those that do not completely resolve) is, to some extent, dependent on other features. For example, if an imperfect cadence ends on a long note and is followed by a repetition, it is more likely to be seen as a strong phrase end followed by a phrase start than if it continues straight on with a the same melody with no change, repetition or gap. Usually the definition and identification of a cadence is independent of temporal information (cadences can be long or short) but the metrical (and temporal) information does affect the subtype and strength of the cadence (and therefore its importance for a phrase end or the importance of the phrase within the larger context).[124]

Harmonic progressions rely on a 'harmony' being present. This does not exclude monophonic music where there is implied harmony. In such cases, however, there is room for ambiguity and it may be that if the monophonic music were too ambiguous (or too monotonic), this feature would not be useful. These

[124] At a more basic level, harmonic progressions are also among the components contributing to metrical structure.

possibilities are also present in polyphonic music. There are cases in which the harmonic information is too monotonic or too complex and where other features are relied on instead. In the monotonic music the harmonic information does seem to generate expectation for change ("this cannot go on for ever") and the change may signal the new start (such as in the Mozart Aria). When the harmonic information is very complex (modulating to or cadencing quickly on distant keys), there are still cases where expectations are generated but when the 'resolution' is so unexpected that this feature becomes difficult to recognise and others may be relied on instead (Bach Passion).

The harmonic characteristics of the phrase not only help define its boundaries and induce the expectations thereof, but also their internal structure, the relation between consecutive phrases and the relative harmonic (and structural) importance of the phrase within a movement (or at least a section).

12.2.4.1.2 Voice-leading – explicit and implicit

Voice-leading is closely related to harmonic progressions such as cadences and to the pitch 'phrase arc' (see for example, Huron 1996, and also chapter 1).[125] The most common (and strong) voice-leading feature of a phrase end is the attainment of $\hat{1}$ in the melody part. However, it is not just the attainment of $\hat{1}$ but the tonal scale-descent to it, supported by the underlying harmonic structure, which define this feature and which is essential to its effect as an expectation generating one. There are several different lengths of descent ($\hat{8}$-$\hat{7}$-$\hat{6}$-$\hat{5}$-$\hat{4}$-$\hat{3}$-$\hat{2}$-$\hat{1}$, $\hat{5}$-$\hat{4}$-$\hat{3}$-$\hat{2}$-$\hat{1}$, $\hat{3}$-$\hat{2}$-$\hat{1}$, $\hat{5}$-$\hat{4}$-$\hat{3}$-$\hat{2}$//$\hat{3}$-$\hat{2}$-$\hat{1}$ etc.) that can occur at different 'levels' within the melody. This feature has its roots in, and is a simplified interpretation of, some aspects of the Schenkerian analytic approach (Schenker 1906, 1954; 1935, 1979, see also chapter 1). It emphasises a 'descent' in the melody part that is supported by the harmonic structure and can be identified at various levels within the musical fabric. At its most fundamental level, the descents span the length of the piece – with interruptions and repetition that delay resolution but that contribute to the subdivision of a piece. Within phrases the descents are over relatively short distances, near or at the surface of the music and the most important parts of the descents are towards the end ($\hat{3}$-$\hat{1}$ to the local tonic).

The identifications of these harmonically supported descents are at levels that are at ('explicit') or relatively near ('implicit') the 'surface' of the music. The 'explicit' voice-leading descents refer to situations where the melody notes are only those of the descent. These are the simplest to identify, and usually result in the clearest expectations and attainments of phrase endings. The 'implicit' voice-leading descents can be ornamented 'explicit' descents (such as a crotchet scale descent, supported harmonically but embellished by semiquavers) or may be more hidden. In both cases, the descent may be either resolved (reach $\hat{1}$) or begin and be

[125] The voice-leading patterns of pitches considered here are the melodic, or most important, musical lines.

disrupted by either being delayed or discarded (such as, $\hat{5}$-$\hat{4}$-$\hat{3}$-$\hat{2}$//$\hat{3}$-$\hat{2}$-$\hat{1}$). As with the cadential progressions, the resolved descents are more a mark for a phrase end, while the effect of actual phrase end with the disrupted descents depends on the other features. The disruption of descents affects the expectation of phrase end and hence the phrase end perception.

Being related so closely to the harmonic structure of the piece, it may seem that either one or the other of the two features are redundant. However, as described above, it is possible to have a perfect cadence where the voice-leading is delayed or not present and *vice-versa*. The relationship between the two is not always the same, and different relationships lead to different types and strengths of phrase-ends. Furthermore, there are some musical styles where one or the other of the cues is not employed such as, in Renaissance music, where the most important relationships are usually described as linear (horizontal rather than vertical or chordal). These voice-leading patterns may be seen as closely related to pitch arcs described by, for example, Huron (1996).

12.2.4.1.3 Change in rate of (harmonic) rhythm

The main relation between voice-leading and cadences and temporal information is that the effect of the former is stronger if the notes or chords occur on strong metrical positions. However, there are also other connections. For example, the expectation for a phrase end is also triggered through an increase in harmonic rhythm. For example, if there is one harmony per bar for several bars and then an increase in harmonic rhythm to a harmony per beat, there may be an expectation that the end of the phrase is approaching. The last harmony of the phrase may (or may not) again be relatively long and the new phrase may then begin with, and be signalled by, the slower harmonic rhythm (as in the Brahms, bars 1–3, 4, 5–7, and 8).

This feature most often contributes to an expectation for a phrase end and may instantaneously return to a slower (or even the original) rhythm at the phrase start, a feature that would probably be perceived in retrospect. The importance of this feature is in the expectation generating part of the phrase rather than the boundary itself.

On one level, this feature contradicts previous features (and the Gestalt principles); previously, a change has indicated a possible phrase boundary, a new phrase. Here, on the other hand, like some changes in texture, the change indicates preparation for a phrase end – the change is in the middle (or towards the end) of the phrase. However, this reinforces the idea that once these expectation features have begun, there has already been an affect on the phrase (the corner has been turned), whether or not the phrase ends where expected, is less relevant.

12.2.4.2 Short term, expectation generating features: delaying suspensions, long notes

The above expectation generating features have mainly been concerned with relatively long-term connections. However, there are some further features enhancing expectation generation that occur usually within cadences and that are much more immediate.

One such feature is the suspension. This delays the resolution of one or more of the parts to a later position than the initial resolution, increasing tension and maintaining the expectation for a resolution. There are also prolongation features (such as repetition of notes/chords or long notes) that emphasise and lengthen a resolution once it has occurred. In both cases one of the features used is long-notes. In the latter, this follows the 'rule' that long notes occur at phrase boundaries. In the former, however, the long note can occur before the end of the phrase, contradicting the rule usually used in Gestalt-based approaches (for other discussions of short term expectation generation see Narmour 1990; 1992, and chapter 1).

12.2.4.3 Summary: Expectation Generation in Musical Features

The different expectation generating features can be seen to fall into four categories. Those whose most important rôle is to:
1) Lead to and reach the end of a phrase (e.g. resolved cadences),
2) Lead to the expectation of the end of a phrase, and then the distinction that a new phrase has begun (e.g. changes in rate of harmonic rhythm),
3) Increase the expectation of the end of a phrase (e.g. short-term lengthenings),
4) Generate the expectation of the end of a phrase, thus changing the character of the phrase but the expectation is not necessarily fulfilled (e.g. thwarted cadences).

As with the other features, for the expectation of the phrase end and the phrase end and start to be strongly displayed, other features (confirmatory or instantaneous) are necessary. However, the study of the case-study pieces shows that, unlike with the other features, these features are almost always completely reliable markers for phrase end expectation and phrase boundaries – whenever there is a cadence, or a resolved voice-leading progression, there is also a phrase boundary of some sort. The exceptions are some of the short-term features, which rely on long-notes.

Furthermore, these expectation features help to describe the areas where phrase-end expectation begins, those where phrases are prolonged and those where phrase-end expectations are thwarted. Using these features, it is possible to move away from the boundary identification and begin to describe features that contribute during the phrase. Moreover, it seems that the exact location of the phrase end or start is sometimes secondary to the identification of the expectation of the end of the phrase. The completed phrase is less important than the phrase expectation.

12.2.5 Standard, template, ideal, expected phrase lengths

Presence	**Instantaneous**	'Template' measure through: - Number of Notes - Clock Time - Number of Bars - Structural or Motivic characteristics
Impact	*Instantaneous/Predictive*	
Result	*Template,* *Boundary or relation to template boundary*	

Intuitively it seems that short-term memory, longer-term memory making connections, and physical constraints (e.g. feeling of breath) must play an important rôle in phrase perception and that it should limit the maximal phrase length. There are a number of candidate features that may contribute to this.

12.2.5.1 Length according to Number of Notes

Having an expected phrase length is an idea used by Temperley (2001). He sets the expected or optimal phrase length by the number of notes. For his corpus he finds that an 8-note phrase length is the most suitable and penalises phrases that are longer or shorter according to the amount they differ from that length. This restriction is simplest to define for monophonic music and if all notes of the melody are regarded as equally important. However, in more complex music where there are notes that are more or less important (especially in situations where some notes can be seen as elaborations or embellishments) the usefulness of note counting becomes more difficult to see. For example, looking at the opening of the Mozart Sonata, bars 0^3 – 2^2 are repeated and embellished in bars 4^3 – 6^2. The number of beats covered is almost identical (the second section is a quaver shorter), the harmonic structure is identical, the underlying melody is the same and the response of listeners and performers is the same. However, in the first section there are 16 notes in the melody and in the second there are 23. With such a variation within a piece, where the music is based on the same theme and harmonic structure, assessing an optimal number of notes (or even range) is not only difficult but also seems to hide the similarity between these two phrases. Having difficulty in looking at just one piece (and therefore maybe setting optimal numbers according to style/piece) it may be more useful to look at more pieces and take an 'average' or range. However, this seems to increase the problem. Analysis, for example of the case-study pieces shows that an eight-note unit is in most cases shorter than many of the phrases identified. Limiting the number to a specific one is overly restrictive for these pieces. Even assigning a range is difficult. Most importantly, counting each note implies that each note is treated as being equally important. Musically, this contradicts the idea of elaboration and, from the perspective of the listener, contradicts other ideas of reduction in perception. However, though the idea of counting notes may not be the best approach, the idea of an expected phrase length still seems to be relevant from three perspectives:

1. An upper limit on a possible phrase length seems intuitive in terms of our ability to remember information. There are however, other limitations on the amount of information we remember including the density of information (elaboration) and how well it can be 'simplified'.
2. Upper and lower limits on a possible phrase length also seems intuitive in terms of the physical restrictions of breath (internal/external).
3. There may be some kind of temporal template which we expect phrases to follow, especially once the first phrases have been played.

12.2.5.2 Length according to Clock Time

There may be both physical (e.g. breath) and psychological (mainly memory-related) upper and (even lower) limits on phrase length. In theory it should be relatively straightforward to ascertain what these limits are through a combination of experiment and study of pieces. This would be particularly simple if phrasing in music were to be treated like any other medium where experiments have shown that the 'present' or 'nowness' is limited to a maximum of 3 seconds and is an attractive idea as it seems to be so generalizable, often referred to as the Psychological Present (Wittmann and Pöppel 1999, p. 842, see also chapter 1). However, there are two problems. As Wittmann and Pöppel (1999) point out

> 'it is proposed that temporal integration up to 3 seconds is a general principle of the neurocognitive machinery.
> This integration is automatic and presmantic**,** that is, the temporal limit is not determined by what is processed, but by intrinsic time constants.'
> (Wittmann and Pöppel 1999, p. 842)

Placing phrase perception in the presemantic realm implies that there is little 'processing'. Moreover, in terms of the analyses and experiments carried out in this study almost none of the segments identified as phrases in this study (and in other studies, such as, Deliège 1998; Temperley 2001, chapter 8) fall within this upper limit. For example, both analysis and experiment suggest that the upper limit of the range may be 11 seconds. In the Mozart Sonata the longest phrase is 17 sec and in the Wagner it is 24 sec. In the Wagner there is a very large standard deviation, suggesting that extremely long phrases cannot be sustained or repeated many times. It remains questionable whether there is an ideal (or general) length of phrase in seconds, or a range of reasonable phrase lengths, what that ideal is, and how it can be measured.

It may be that the internal punctuation marks are more relevant for this and it seems that the 'temporal integration…subjective present … feeling of nowness, … singular single states of being conscious' Wittmann and Pöppel (1999, p. 842) refer to smaller groupings within music than the phrases that have been identified by theorists and in the studies discussed in earlier chapters.

The study of performances and responses to them (chapters 3-11) indicate, the same 'phrases' can be played at different tempi and therefore last different lengths: when they are played more slowly and last longer, more internal subdivisions are identified. On the other hand, when they are played more quickly, there is much less emphasis on the shorter phrases and more on the longer. To some extent, this supports the idea that there is some kind of upper limit for length of phrase and if it gets too long, the shorter phrases are identified. However, as the musical examples show, the main 'longer' phrases are still those most identified, even in the slower rendition, indicating that the relationship is not completely straightforward (chapters 1, 4, 7 and 11).

12.2.5.3 Length according to Number of Bars

The number of bars in a phrase has been often taken as part of its defining features. For example, Four + Four (or Eight + Eight) bar structures are often given as basic examples of 'a phrase'. This kind of structure, along with its extensions is advocated in composition manuals (such as Koch's 1787, 1983) and was mentioned by some participants in the current study. In the pieces studied, the first beat of four bar sections coincided with phrase starts or ends in more cases than not. Furthermore, there are several phrase starts that occur in the 'expected' bars.

Furthermore, once a certain length has occurred it is often described as being expected to be repeated (e.g. Koch 1787, 1983). In the pieces analysed in this study, phrases identified were both of equal and unequal lengths. Moreover, Koch, for example, teaches that phrases should begin in the same part of the bar (Koch 1787, 1983). In the test pieces analysed here, the phrases often (but not always) start in the same part of the bar as the opening phrase or on the equivalent position in the half-bar (e.g. both Bach pieces). There may be expectations for a phrase boundary on a particular beat, which is then delayed by, for example, embellishments or suspensions and therefore automatically searching for positions in the bar as phrase boundaries would not identify all phrase boundaries but could be informative. In the pieces studied here, the phrases identified varied in lengths from 1 bar (or less) to approximately 8 bars. An automatic search for phrase boundaries every 4 or 8 bars would probably identify some but others would be missed.

Setting a usefully predictive template is therefore difficult but could be informative. It reflects a possible expectation for symmetry and organisation, especially during the opening of the movement. However, these basic expectations can be thwarted so this method would not always identify the correct positions. In a large corpus, it may be a useful measure.

So far this discussion has treated bars as measuring units. However, they represent more than this. The relation between the metrical and phrase structures affects the character of the phrase; phrases that start and end on different beats have different momentum, strength and finality. This feature is therefore more useful in characterising phrases than identifying phrase boundaries.

12.2.5.4 Structural or motivic information

It seems that there are several factors that override the importance of absolute phrase length. Even the theorists that advocate phrase lengths in bars, treat these as 'basic' phrase-lengths, even though the actual length of the phrase may differ (through for example, extension or elision). There are other ways of calculating the length of a phrase. For example, according to Rothstein, underlying structural principal tones may determine phrase length and are important to the extent that if contradictory to the usually favoured duple metrical phrase structure, they may override it. This is one of the most significant ways of constructing a non-duple phrase. For example, in Schubert's 'Wanderers Nachtlied' (D. 768) a phrase of five half bars corresponds to the succession of five principle tones in the melody rather than metrical characteristics (Rothstein, 1989, 1981, Schenker, 1935, 1975, pp. 120-125, and Schachter, 1987).

The results of this study indicate that within a piece, the internal structure (melodic, harmonic and rhythmic) of each phrase, the relation between phrases, the identifiable elaboration within phrases, are the features that contribute to the determination of phrase length and that could be causing listeners to expect phrases of a certain length.

12.2.5.5 How these measures may be used

Clock time and number of bars (in terms of both total length and template), may be most useful as a check or filter to eliminate the most unreasonable options and may provide information as to where the phrase end may be expected. Phrase length seems to be an important factor and there are different ways representing this length in a musically meaningful way, taking account of structure, relation with previous phrases, metrical structure and overall tempo of piece or performance.

Phrase length relates to the similarity or repetition features and the predictive and confirmatory features. However, at least for some of the ways of calculating the length, this 'feature' would need information from some of the others. Rather than giving absolute positions for phrase ends or starts, phrase length contributes to expectation, both in terms of length of individual phrases (they are expected to be "not too long" and "not too short"), and in terms of regularity in length or ratio. In addition, a phrase start or end at a particular metrical position may increase expectation for phrase starts or end in the same position later in the piece.

12.2.6. Performance features

Presence	*Instantaneous/Gradual*	Tempo change – - overall tempo, - long-term tempo change (and rubato) - short term tempo change Dynamics - gradual/long-term dynamic change
Impact	*Predictive*	
Result	*Expectation Generating*	

Presence	*Instantaneous*	Breath
Impact	*Instantaneous*	
Result	*Gap*	

The results of this study indicate that performance features (some of which are also expectation generating) provided by the performer (on the basis of the musical features) often emphasise the musical features, influence our perception of the phrases and, in difficult passages, may clarify. For example, if musical features indicate thwarted phrase ends followed by a clearer phrase end, the performer may riterdando more towards the final phrase end than towards the preceding ones. Here tempo and dynamic change were primarily investigated.

Tempo change is often taken to indicate phrasing (for example Todd, 1985, Clarke, 1988, Repp, 1990, also chapters 1 and 7). This study also shows that phrases often seem to be reflected in, and accentuated by, tempo change. However, other structures, such as metrical structure, are also reflected in tempo change (chapter 7). For each type of musical element the manner in which the tempo is changed may be different, and so the same performance feature may highlight different structural aspects and the clarity of this distinction varies. Even in terms of phrasing, the study indicated that there could be least two different types of tempo change: long-term (a new phrase has already begun), and short-term (in preparation for phrase boundary).[126] As with tempo change, it is possible to distinguish between long-term change (new phrase) and short-term fluctuation (preparation for phrase boundary) in intensity.[127]

One of the key characteristics of performance features, and one of the reasons that we keep listening to the same pieces performed over and over again, is that they can (and often do) vary among performers.[128] Different performances of the same piece may emphasise different elements of the structure (chapters 3, 7, 10

[126] There were no pauses in the case-study scores. There were some in the test pieces (discussed in chapter 15).
[127] Lerdahl and Jackendoff (1987) include Dynamics (and Articulation) within their GPR 3.
[128] Though they vary more in Romantic music than in Baroque and Classical. See, for example, the differences between the different performances of the Mozart Aria and the Brahms.

and 11). In this study, such performance features are treated as in some way representing the performers' interpretation of the phrase structure rather than 'only' structure-generating features.

12.3 Feature Categories Revisited

The features were here discussed according to their feature categories (section 12.2). Each category has several characteristics, each reflecting a different aspect of the features and their effect on the phrase structure. There are two temporal characteristics, one relating to the feature's 'presence' and the other to its 'impact' and function or 'result'.

12.3.1 The temporal characteristics (scope)

As introduced above and in chapter 11, the scope of effect of the features varies in terms of the length of presence of the features, the area that they affect, and the reaction of a listener. Each scope category has a different kind of impact on the phrase perception. Most of the features belong to only one category but some features can belong to different 'impact' categories.

12.3.2 Instantaneous Feature:

12.3.2.1 Instantaneous Feature – Instantaneous Perception

The feature group with the shortest scope is when the feature's presence and perception are almost instantaneous (for example, pitch jumps, long notes and rests). The new phrase is expected to, and perceived as, beginning on the note after the pitch jump, rest or (some) long notes.[129] There is no preparation inherently involved in this feature category.

12.3.2.2 Instantaneous Feature – more than one 'impact' category

Notes falling at the beginning of a bar may also be included in the previous category. However, under some circumstances, the bar line (i.e. relative metrical strength of the start of the bar) and four-bar template (i.e. relative metrical strength of longer sections) may be perceived instantaneously but its scope of effect may be wider; the first metrical stress of the phrase may be expected at a certain position and though the phrase begins a beat or two early or late, it is connected to the metrical position. These features are therefore here referred to as affecting an area around a position.

[129] Change in note-length may also be in the instantaneous occurring but retrospectively perceived category.

12.3.2.3 Instantaneous Feature – Retrospective Perception

Changes and repetitions occur more or less immediately (their start position is usually one note or rest) and may be begin to be perceived almost immediately. However, a definite decision requires accumulation of evidence. Once that decision has been made, the beginning of the new phrase is described as having occurred at the position where the change occurred. In terms of impact, this is considered a retrospective feature group and includes exact and inexact repetition, changes in motive, texture, and long-term note-length, (and tempo and intensity) changes.

12.3.3 Gradual (Predictive) Feature

12.3.3.1 Gradual Feature – Predictive and Retrospective Perception

There are features that prepare the phrase end (and sometimes the next phrase start). These include harmonic progression (especially cadences) and voice-leading, change in rate of harmonic rhythm, repetitions of closing motives, and some long-notes and changes in tempo and dynamics. In such cases, the feature is perceived gradually and an expectation for certain patterns leading to the end of a phrase continues or develops throughout the feature's presence. Reactions to these features are, therefore, both retrospective and predictive. The resulting 'prediction' of the position of a phrase end (and following start) may be on a specific beat or over a general area. When the resolution occurs, its perception can be instantaneous or primarily retrospective.

Retrospective and Instantaneous features following other features can often act as confirmatory features, reinforcing a phrase boundary that has already been expected or detected and often clarifying the position of the phrase start. It seems to be for this reason that the combination of features from different feature categories is so strong.

12.3.4 Templates

There seem to be several possible weak expectations for the phrase to match a 'template', in terms of clock time (an upper limit of length), bars (an even number of four, eight or two), 'symmetry' (the lengths to be the same all phrases) or metrical position in the bar (the same metrical position in all phrases). This would fall under the instantaneous position identification and would be predictive. However, this often seems weakly related to what occurs (though this in itself may be informative).

12.3.5 Feature Combinations

The more features that support each other, especially if from different feature categories, the clearer the phrase boundary and the more likely there is to be a high response among participants. This approach implies an additive model. However, there are some features (such as the harmonic ones) of which only a small number are necessary for clear phrase identification.

12.3.6 Scope – the presence and representation of the features

The presence and impact of some features are limited to beats or notes, while those of others extend over a range of notes or beats, and some are a combination of the two. The scope of perception and effect of these features may change under different circumstances (i.e. depending on the combination and position of features). As discussed in chapter 11, in the graphic representation in this study the feature is represented at the first possible phrase boundary note suggested by the feature.

In some, (computational) approaches instantaneous features have been concentrated on (Ferrand et al. 2002, Bod, 2001). In others, instantaneous and retrospective features have both been studied (Lerdahl and Jackendoff, 1987, chapters 1 and 8). The results of this study indicate that one of the most important parts of phrase (and music) perception are the predictive, expectation (and interest) generating features, and consideration of types is important.

12.4. Summary

The above discussion assessed each feature in detail and attempted to show the usefulness and limitations of each for phrase perception. The features (along with their different gradations) can be grouped according to their scope and function. In the following section, the music will be looked at from the other direction with a discussion of the different phrase-types and structures (chapter 13). These are both summarised and formalised in a rule base (chapter 14).

Chapter 13

What are musical phrases?
The arrival

13.1 Introduction

The listening, annotation and performance studies and the exploration of musicological, music-psychological and computational literature have shown that the term 'phrase' often refers to a musical unit with general possible and various characteristics and/or functions (sections 13.2-3). Within this, there are a variety of different phrase types, each with its own characteristics and functions (or rôles), which will be discussed here (section 13.4). The various internal characteristics of the phrase and how these can differ among circumstances, contribute to the differences between phrase types. The identification of these characteristics is dependent on different musical features and feature combinations. Some of the different musical features explored that can dominate in the different phrase-type categories are mentioned again here in the context of the phrase types (section 13.5). The relation between phrase and metrical structure is then briefly discussed (section 13.6). This is followed by a more over-arching view describing different possible relationships between phrases (section 13.7). This relates to the question of phrase length – the usefulness of regarding phrase-length as a defining element of what a phrase end is (section 13.8).

13.2 General Structural Aspects

As a basic template, a phrase is a component of a piece that should have a beginning and an end,[130] and is distinct from, but related to, the preceding and following phrases. However, the basic structure can be played against, by for example weakening the beginning and delaying or interrupting the end. Moreover, the defining structural constituents of a particular phrase, those that mean that it is identified as a phrase and those that define its 'function', may not be the beginning, end or either but that: 1) there is an expectation for these elements, 2) there is a third key element which is the part in which expectations for phrase closure is initiated within the phrase, and 3) different elements may be more explicit or strong whether they are the beginning, the end, the expectation for the end of the phrase or any combination of the three, thus changing the character or types of phrase. The beginning, end, and expectation of the phrase end may be of different 'strengths' and the stronger components help to define the phrase types. The different strength of ends or expectation for ends leads to a stronger or weaker need for a continuation. A strong beginning may (in retrospect) strengthen a preceding weaker end and vice versa.

One possible description of the phrase, therefore, is that it is aiming or searching for a strong close.[131] However, as the listener, annotation, and performance studies have shown, the lack of a strong close does not mean that a phrase will not be perceived. That the aim for a strong close is frustrated also seems to be a possible component of phrases. On the way to that close, therefore, there can be several weaker closes that do not eliminate the expectation for the strong close but may affect the urgency for one. It is one of the defining characteristics of the later nineteenth century tonal music that it pushes the delay to the extreme. But as seen in the study-pieces, these delays are not limited to the music of that era. There is a gradation of 'strong' closes within a piece which is one of the elements that determines the relationship between phrases and helps to structure the music for performer, listener and analyst. Viewed in this way, phrase structure (both within the phrase and over the course of the piece) can be regarded as having relative strengths.

13.3 General Function of phrases

It seems that the phrase is one way in which the structure (or structures) of the western classical music studied here is expressed. The phrase may help the listener, performer and analyst in, for example, following the progress of the music, identifying segments (to be remembered) while the piece is proceeding, and

---

[130] This view is implicit in any use of the term of which I am aware.

[131] See for example, Rothstein (1989) who, among others, interprets this close tonally ('tonal close') and chapters 1, 9-12.

memorising the music in the long term. Research has indicated that phrases are indeed used in, for example, practice for performance (Chaffin and Imreh, 2002).

13.4 Phrase-Type Categories

It seems that some general conclusions may be drawn regarding the different possible phrase-type categories and the features that contribute to them. Different phrase parts or phrase part combinations may be important in different contexts. As discussed in chapter 11.3.2, a number of broad phrase-type categories may be identified. In relating these to the musical features (chapter 12), the associated features and feature-categories are briefly mentioned.

In what follows it should be borne in mind that, as the listeners and performers and even analysts studies show, the phrases highlighted and/or identified may be different for different people, or the same people on different hearings.[132] This may be for numerous reasons including the attention of the listener, performer or analyst and may include where the attention of the listener, performer or analyst is directed both in terms of the type of features they pick-up on and the level of completeness they regard as necessary in their phrase definition. It may also depend on the musical experience or familiarity with the piece.

A. End-heavy, expectation dominated, goal directed, left branching phrase types

This group refers to phrases whose most important aspects are that they lead towards (and sometimes attain in a closed or resolved manner) the end. In the more straightforward cases, where there is only one expectation and one resolution, the end of the phrase is prepared (expected) and is 'completely resolved'. Here 'completely resolved' is relative, referring to complete resolution onto, for example, the local tonic in both the bass and melody parts (for example Mozart Sonata, bb. 7-8). Alternatively, there may be temporary 'partial' (transient) resolutions or just expectations for resolutions that do not resolve at all (for example Bach Passion, bb. 3-4). These types of phrases are here described as end-heavy to capture the idea that, in phrase-structure terms, the most important information is leading towards and at the end of the phrase.

Phrases within this category are likely to have expectation generating features such as, cadential progressions, voice-leading progression (for both of which the level of resolution will affect the extent to which the expectations have been fulfilled) and increase in rate of harmonic rhythm. If they resolve clearly, the phrases may have clear instantaneous features (such as gap features including pitch jumps and long notes).

[132] Or even, as the results show, people are often aware of more than one interpretation at a time.

B. Start-heavy, statement or confirmation dominated, right branching, phrase types

In contrast, for front-heavy phrases, the beginning of the phrase is the part with the most important information. The ending of the previous phrase does not necessarily have to have been prepared or explicit for the new phrase to be felt. Instead the phrase start is the distinctive part. Unlike the goal directed phrases, which drive forwards, this type of phrase can be broadly described as making a 'statement' and may be a likely candidate for thematic importance and may therefore be important to remember. In fact such 'statements' are often repeated (with or without modification, both immediately and after some time), increasing the listeners' chances of remembering the material and 'appreciating' its importance (such as in the Brahms and Mozart Sonata).

Phrases within this category are likely to have change or repetition features such as change in motive or change in texture or different types of repetition (including exact or inexact repetition). Again, the type of repetition may influence the strength of a phrase start.

C. Instantaneous division, breaths or gaps, phrase boundaries

In some cases, phrases are most clearly distinguished through their gap features including rests and long notes. When alone, there is little preparation or confirmation and instead the boundary itself is the distinctive part.

A phrase can have all three elements: the clear phrase start and the goal directed, resolved part and a clear 'boundary' preceding and/or following it. These would be examples of the clearest phrases (for example, Mozart Aria, bb. 5-8). However, as has been suggested above, there are also a number of possibilities within these two general phrase types with most of the variety being within the first (goal-directed) group.

More complex phrase types

These components can be combined in several additional ways:

f) Delayed phrase end (following preparation/expectation)

This type falls within the end-heavy, goal directed category. However, here the resolution following the preparation for, and generated expectation of, it is not immediately achieved. In some cases an expectation (continuation of the first one or a new one) is (re-)ignited after the first expectation is not resolved. After an unfulfilled phrase-part, the expectation for a resolving phrase continues. The perceived phrase status of the unresolved phrase-part depends on the musical features and the strength of temporary resolution they suggest, the experience of the listener, the familiarity with the piece, and the 'view' on the piece; a general or retrospective 'view', where the problems of online memory are less important and long range connections become more important, or a moment-by-moment one where the relative importance is reversed. For example, the attempt at a resolution is sometimes, if seen in retrospect or from a general or retrospective view, may be regarded as more of a punctuation within a phrase (for example, Brahms, bb. 6-9).

Delayed phrase ends which following preparation/expectation features (such as cadential and voice-leading progressions) but they do not resolve and do not have the clear instantaneous features.

g) Equally important adjacent sub-units/phrases – related to unfulfilled expectation

One example of the punctuated phrases occurs when there are equally important sub-units that rather than being seen only as punctuated segments of the phrase, have a symmetry and structure that brings them closer to the structure of a phrase but where the strong resolution still only arrives after more than one such unit. These are often symmetrical and can form an internal antecedent-consequent relation (for example, Mozart Sonata, bb. 1-4). This is the 'ideal' subdivision for Lerdahl and Jackendoff, who treat phrasing as a middle level within grouping levels, for whom this structure is part of the definition of what a phrase is (both its internal structure and the relation between consecutive phrases).[133] In the current study however, this relation is regarded as just one option.

In terms of equally important sub-units (phrases), taking for example the antecedent-consequent pair the expectations set up in the first unit are fulfilled only in the second. For example, the expectation generating features of cadential progressions and voice-leading are begun in the first part and only resolved at the end of the second.

h) Sequence

Another such case is that of a sequence where often each subsection of the sequence is self-contained but is part of a complete phrase (the end of which may not be clear).

In sequences, the repetition feature returns but this time in shorter units and more repetitions of the same material. In terms of the harmonic features, the repeated units are not completely resolved at the end of each repetition but may be resolved at the end of the series of repetitions.

---

[133] GPR5 Symmetry (Lerdahl and Jackendoff 1983, p. 346), see chapter 8.

i) Elided phrase or overlapping phrase boundaries

In many descriptions of phrases the most important element seems to be the phrase 'boundary'.[134] One of the possible implications of this view is that there is not necessarily a clear distinction between the phrase end and the phrase start. As these are different parts that relate to each other differently under different circumstances, clear distinction between a phrase start and end is, however, essential. This is even the case when the phrase start and phrase end occur simultaneously (usually called an elided phrase).

In 'simple' elided phrases, the phrase end and start coincide on the same note. For example, the arrival on the tonic at the end of cadential and voice-leading progressions is also the start of a repetition. However, in more complex cases there may be cases where the overlap of phrase end and phrase start is over a longer duration in which the previous phrase ends after the new one has begun (for example Brahms, bars 5 and 9). Such cases are most often identified as overlapping by analysts.[135] However, in terms of the listener it seems that either the phrase end or phrase start will dominate and therefore shift the perception of the other to match. So for example, if there is a strong phrase start, followed by a weaker close of the previous phrase, the strong start will dominate and maybe acting as an interruption will shift the perception of the phrase end to the position of the phrase start, or even just before.[136] Even here then, a clear distinction between phrase starts and phrase ends seems important.

Such elided phrases are seen in a different light if they are seen against a template which separates phrase ends and phrase starts, emphasising that, in an elided phrase, two separate parts (with different functions) are coinciding.[137]

j) Ambiguous phrase areas

The previous case of overlapping phrases is also an example of the ambiguous areas. Here ambiguous can refer to two scenarios; 1) the possible phrase start and phrase end positions do not relate in the usual way, such as in the more complex situation described in d) above, or 2) there are different possible positions for

---

[134] See for example, Lerdahl and Jackendoff (1983) Temperley, (2001), Bod (2002). Though Lerdahl and Jackendoff often concentrate on 'grouping' boundaries rather than phrase boundaries (chapters 1 and 8).

[135] For example, discussions of the Brahms by Burnett (1972) and Dunsby (1981), chapter 10.6.

[136] For example, the first three semiquavers of Brahms bars 5 and 9, and the discussion by analysts Burnett (1972), Dunsby (1981) pointing out that the harmonic resolution is only achieved on the third semiquaver of the bars (onto b) while the thematic beginning starts on the first beat of the bar (chapter 10.6).

[137] The term 'elided phrase' is used by, for example, Koch (1787, 1983) and Lerdahl and Jackendoff (1983) (chapters 8 and 9).

phrase ends and phrase starts suggested by the musical features. There are three possible responses to these: i) The analyst can identify these ambiguities and 'celebrate' them. ii) The performer identifies them and picks one possible structure to highlight (if it is too complex, the performer cannot). iii) The listener, if presented with a 'straight'/'unclarifying' rendition identifies them and either chooses one possibility or gets confused. If presented with a clarifying performance one of the options is made most explicit thus reducing or removing the ambiguity for the listener.

f) Internal punctuation positions and online-listening

In all these cases, the nature of the structure of the phrase (unfulfilled expectation, internal symmetry, ambiguity) has a crucial effect on the phrase perception and, in all but the ambiguous situation, on the possibility of having rather long complete phrases which have internal punctuation. From the experimental results these punctuation points are perceived both by listeners and performers. Often keys are pressed at least in one listening with the expected end such as at the place where the phrase should have resolved or at the place where the music is 'interrupted' with a new start. After listening, participants in the experiments often commented that, in retrospect they would have excluded these midway punctuations (but they often did not in subsequent hearings). Here a distinction can be drawn between the 'online', real-time, immediate perception of the phrase, where the individual points of punctuation are used on the way to the stronger resolutions, and the retrospective perception of the phrase (or when the piece is known better) where the longer, resolved phrases are the ones that are foremost in the perception (the kind of response also depends on the musical experience of the listener).

One of the results of the reassessment is that there is potential for constantly changing view of the phrase structure which is one of the elements that may keep us listening. To some extent, we expect a regular structure but given features that contradict this, we are ready to re-assess. Once given the 'true' phrase end we may have to reassess earlier conclusions or they may no longer be relevant.[138]

13.5 Internal Structure of phrase

As already discussed in section 13.4 (phrase-type categories), different parts of the phrase seem to have importance in different contexts. In this study, it is suggested that certain parts of the phrase play a particularly important rôle.[139] Overall, the

---

[138] See also for example the discussion in Jones (1992, chapter 1).

[139] In some ways this is similar to other approaches, such as the more music-analytic one taken by Rothstein (1989), or gestalt-based approaches which end up concentrating on boundaries (e.g. Lerdahl and Jackendoff 1983). However, these approaches seem to concentrate on boundaries to the exclusion of what leads to, and follows, them. In this approach, on the other hand, importance is placed not only on the boundaries, but parts of the phrase and the features that lead to, and follow, them. This approach also attempts to include more features than are usually included in other studies, which, for example,

results of the studies carried out here suggest that there are five parts of the phrase that seem most relevant which will be discussed in turn. To enable reference in discussing the rule base and algorithm (chapter 14), the 'states' representing each phrase part are also presented here.

- **Phrase Start** (S_1)
- **Beginning of the expectation of the end** (S_2)
- **Beginning of the end** (S_2)
- **Arrival at the resolution and prolongation of the end** (S_3 and S_4)
- **End of the end (end of the resolution)** (S_5)
- **Phrase gap** (S_6)

- **Phrase Start** (S_1)

The phrase start is one of the clearer parts of the phrase. It is clearest at the beginning of the piece which gives at least some expectation for what some of the following phrase starts may be like. Moreover, even when phrase starts are not directly related to the material at the start of the piece, clear features often signal the beginning of the phrase. The phrase start can establish the thematic and tonal centre of the phrase and provides the 'starting point'.

One of the difficulties with phrase start identification in the studies carried out here occurred when a phrase start did not coincide with a metrically strong beat. In such cases, responses tend to range over the area from the weak to the strong beat. It seems that the problem is on one hand related to our normal response patterns in which we respond more quickly down beats than upbeats and on the other, to the relative importance of strong and weak beats. However, for most participants and analysts, the weak beat was certainly considered part of the phrase. Another interpretation could be that listeners were responding metrical structure and not phrase structure. However, the responses were rarely systematically related to metrical structure, instead, as discussed below, the metrical position of the different parts of the phrase instead impacts on the character of the phrase.

- **The beginning of the expectation of the end** (S_2)

In terms of identifying a clear position, this is the least exact of the five phrase parts with the position being, in some cases, better described as an area – a gradual realisation that there is a "direction" and there will be an end. This can differ between listenings and listeners. Of course, it is most crucial when the work is unknown to the listener. Once the work is better known, the phrase beginnings and endings are better known and the degree of unknowing or blind expectation is decreased. However, as discussed in Juslin and Sloboda (2001, p. 92), even when a

---

concentrate mainly on harmonic aspects (Rothstein 1989), or gestalt approaches (such as Lerdahl and Jackendoff 1983) (chapters 1, 8 and 9).

piece is well known, some kinds of expectations may still have a function (see also Jackendoff, 1992, and chapter 1).

The beginning of the end (S_2)

This is more restricted than 'the beginning of the expectation of the end' and arrives either with or after it. The beginning of the end is identified, for example, once the beginning of the cadence has been identified, the final melodic descent and/or the beginning of the last rhythmic section. Here the level of expectation is different – the expectation is clearer and the end can be nearer. Once the beginning of the end has been identified, again the character changes. Even if the end does not arrive where expected, there has been enough of a change for later music to be considered somewhat differently to what has come before. In order to avoid repetition of very similar information, the rule base and algorithm of chapter 14, consider these two phrase parts together.

- **Arrival at the resolution and prolongation of the end** (S_3 and S_4)

Within the context of a 'basic' phrase, the resolution (harmonic, melodic, rhythmic) is the same as the end of the phrase. The end of the phrase has arrived and the next phrase, or the end of the piece, is expected to follow immediately.

In many cases, especially in phrases that occur at the end of larger sections or pieces, the resolution may be followed by a continuation of the same phrase in the form of an extension[140] or a rest (especially in music with more than one part, for example the end of the Bach Passion). This extension can be a matter of notes, beats, or bars. To some extent there could be an expectation for the next phrase to start at any position following that resolution, and any 'delay' may have the function of either accentuating the resolution or increasing the expectation and tension for the beginning of the next phrase.

The arrival at the resolution acts as an 'internal punctuation' of the phrase. It is often at these positions that listeners in the current studies lifted their fingers rather than waiting for the actual end. And it is about such positions that they would comment that in retrospect (almost immediately) they realised that they had moved too early and had 'made a mistake'. Interestingly, many repeated this more than once, suggesting that even though in retrospect they thought the phrases were longer, while listening, they reacted to the expectations and resolutions more quickly than remembering what they had learned from the previous hearing. In other words, it seems that, in online processing, the response based on generated schemas based on previous information are winning out over the evidence provided, whilst, in retrospect, the structure suggested by what actually happened is what is kept in the memory.

---

[140] c.f. Koch (1787, 1983) see also chapter 9.

This internal punctuation point may help in the memory and structuring of the phrase. If the section after the resolution is long, it could be perceived as a separate sub-phrase or even phrase depending on its musical features. In the rule base and algorithm of chapter 14, the arrival at the resolution and prolongation of the end are treated separately.

- **End of the end (end of the resolution)** (S_5)

The phrase end is often the goal of the phrase (however temporary in the context of the piece as a whole). When it is successfully reached, the phrase is complete and the phrase structure is usually clear to the listener. When it is not reached, delayed, or less important because one or more of the other phrase parts outweighs its importance, the phrase 'boundaries' are less clear, but the expectations and the crucial structural elements are not necessarily lost.

- **Phrase gap** (S_6)

In some cases there is a clear temporal or intervallic division between two phrases. As will be discussed in chapter 15, this gap can sometimes be filled with a linking section (sections 15.2.1.4, 15.2.2.3 and 15.4.2).

There are certain features that dominate different parts. For example, repetitions and changes in texture and motive indicate phrase starts, the beginning of the expectation of the end is signalled by, for example the beginning of a harmonic progression, the beginning of the end is signified by the more specific cadential progression, the arrival at the resolution can be signified by the arrival of a melodic descent on the local tonic and the end of the end, can be signified by the end of the tonic or, in retrospect, can be signified by gap features such as pitch-jumps or rests. As mentioned above, within these broader phrase parts there can be different strengths of beginning and resolution (depending on the strength and combination of features).

Strength of resolution

There are many different types of resolution that may all be interpreted as phrase ends but each with a different character. For example, the ending may be stronger or weaker, in need of a complementing phrase or a continuation, may suggest an independent phrase, or could or does close the movement or piece. The combination of musical features determines which character the phrase end has. Therefore, it may be possible to distinguish between different phrase types depending on the feature-types and feature combinations. For example, there can be a strong phrase end if there is a resolved cadential progression and root in melody and bass. On the other hand, if the same cadential progression only has the root in the bass and not in the melody, the phrase end is weaker. The resolution is even weaker if the parts are the other way around; the root in the melody and not in the bass. The 'weaker' phrases are one group of phrases that are more likely to be 'in need of a continuation'. Others include those that cadence in the dominant. 'Independent' phrases are those 'statement' phrases with a clear beginning and a strong close (often found at the beginning of pieces). Those that could and do close a piece or movement are also strong closes that finish in the tonic and often elaborate upon it. These are closely related to the different strengths of the musical features, primary, complementary and online, introduced in chapter 11 and developed further in chapter 14.

Relation with unity

If the musical features do not follow the expected development, according to some analysts this has an impact on how we hear subsequent material and, more importantly, we create sense in order to make the music coherent (Kramer 1993, see chapter 9, appendix 9). For others, especially those musicological analysts criticised by Kramer, these passages are ignored in favour of emphasising the overall unity of the piece. However, as we have seen from this study, it seems that the overriding response is not necessarily to immediately create sense in the previous, unexpected (non-unified) material or to ignore it, but to react to it by beginning to generate new expectations. This seems to be done on the underlying assumption that there will be order in the end, and that it may be delayed. With the right level of unexpectedness (disunity), interest is created, with too much, confusion.

These same theorists (such as Kramer, 1993) argue that 1) any coherence is primarily in our heads and that 2) disunity is not less important than unity. This study suggests that, though the search for unity may be 'in our heads' there are musical elements that listeners, performers and analysts all identify and that though disunity is indeed important, it can only be recognised on the bases of 'unity' or a supposition of an underlying structure.

13.6 Relation with other elements, complementary structures: metrical structure

Alongside the 'phrase' structure proposed in this study there are other structural elements that help in performance, analysis and listening. Though they can be regarded as separate and being of different interest to the listener, performer or analyst, they interact with (and sometimes coincide with) phrasing. One of the most studied is metrical structure. Not only is it usually regarded as an important structuring principle on its own, it has an effect on phrase structure. The positioning of the phrase with regard to the metrical structure (up-beat/down-beat) affects the character of the phrase, as does the internal accent structure (i.e. the relationship with the accent structure).[141] For example, if a phrase begins on an upbeat there is often (predictably) a sense in which it is leading to the next downbeat (which is where the next clear harmonic information is usually also

[141] c.f. Koch (1787, 1983) (see also chapter 9).

given). For example, in the Mozart experiments we see that some listeners press a key at the beginning of the phrase (on the weak beat or off beat), and others press the key on or just after the next beat.[142] Similarly, if a phrase ends on a weak part of the bar there is a sense in which the strong beat is delayed – in this way the phrase structure is affecting (maybe even distorting) the perception of the metrical structure.

13.7 Relation between phrases

Phrase boundaries

As mentioned above, phrase ends and phrase starts are often considered together. Although there has been a distinction between phrase ends and phrase starts in this study, this does not mean that the two are independent and separate. Indeed, as discussed already, one often emphasises the other.

From a temporal perspective there are three main relations between phrases: 1) The phrase end is followed by the subsequent phrase start on the next note, 2) The phrase end is on or followed by a rest ('temporal gap') before the next phrase begins, or 3) There is an overlap between the end of one phrase and the beginning of the next (the phrases are elided or overlapping).

These different temporal relationships between phrase end and phrase start can result in different types of phrases and different types of relations between phrases but also different structural interpretations on a more general level. Phrases of type 2) above are the most easily compartmentalised, while two elided phrases may be considered together as one larger unit (with a result that more has to be kept in memory).

Identifying the exact position of the phrase end and the phrase start can be difficult even in what can be considered very simple cases and the problems increase with more complex (and polyphonic) music. For example, should a rest be considered part of a phrase? Does it need to be included to reflect some metrical structure or durational aspect, or does the phrase end with the last sounding note? This dilemma is one that is most likely to be noticed in the experimental where listeners are asked to identify these positions. As indicated by the results of the current studies, it seems that the exact position of the phrase start and phrase end and the moment at which these are identified while listening are less crucial than other aspects of phrase perception. Nevertheless, the click studies (chapter 5) suggested that when phrase ends are clearly on one beat, they

[142] There could be three explanations for this: 1) there are some who do not see the offbeat notes as belonging to the new phrase and intend the strong beat to be the phrase start; 2) We are so trained to beat in time to a beat so that pressing a key on an offbeat is counter-intuitive; 3) There is a sense in which the phrase start becomes bi-polar – the anacrusis is an extension back of the phrase start.

are clearly recognised there. In contrast, when they are prolonged over more beats, this is reflected in the responses.

Relation between consecutive phrases

Relationships between consecutive phrases can be of different kinds such as complementing or contrasting mainly depending on the strength of resolution between phrases: On the basis that each phrase in the end 'completely resolves' (to, for example, the local tonic), harmonic relations of, for example, antecedent and consequent, elaborations, modulations and long-term cadential structures, can be developed between consecutive phrases. These may or may not be explicitly noted by listeners but are part of the formal structures of the piece and may help explain patterns of listening and attention and (maybe) taken into account by performers. These relations can be, for example, harmonic (resolution/change), thematic (repetition/change), textural (continuation/change) etc. Other phrase relations, used by Schubert and other composers, include the 'statement-echo' pairs (the obverse of the antecedent-consequent); both phrases are relatively and equally closed, the second phrase cannot complete the first. The semantic effect is to make the second phrase an amplifying, clarifying, or intensifying echo of the first (Greene, 1970, p. 185, see also chapter 2).

Relation between more distant phrases

On a broader scale, the phrase structures also help "summarise" the music as we listen. It may be that the summarising helps in identifying, and therefore perhaps remembering, more easily the different phrases and their functions. These functions include, 'statement', antecedent, consequent, and closing, repetitions and expectation phrases. This, in turn, enables the identification of longer-term structure and relationships and in generation of longer-term expectations.

Hierarchy

Musical structures are often described as hierarchical (Lerdahl and Jackendoff, 1987, Clarke, 1988 p. 2, and Meyer, 2001, as discussed in section 1.2.9.3). This usually means that parameters are represented as a series of levels, between which relationships of reduction or elaboration operate. In some, the structural levels must fulfil certain requirements, such as being recursive not overlapping. In terms of phrasing, these strict hierarchical rules suggested by Lerdahl and Jackendoff (1987), require that phrases should not overlap or be elided.

A wider view of hierarchical structures, however, may be applicable to phrasing in various ways. One of the main relationships of hierarchical structures, particularly in studies of the perception of natural language, is that of tree structures. One of the main aspects of such a hierarchical structure is that the surface features can be represented as one side of the perceptual process. Information can be summarised or abstracted at different levels, in a way, it is suggested, similar to that by which we process information. Furthermore, the relationship between the surface features and summarised levels are often presented with relationships representing relative importance (dominance, elaboration etc). In terms of the current approach to features and phrasing, it seems it is hierarchical in a number of ways:

1) the surface information is summarised in the form of musical features
2) the features are summarised in terms of their respective phrase parts (including phrase starts and phrase ends).
3) the combinations of phrase parts are summarised in terms of phrase types
4) the combinations of phrase types can be summarised in terms of phrase combinations
5) These phrase combinations of a whole piece could also, potentially, be summarised.

These hierarchical relationships resemble many music analytic approaches (such as Schenkerian approaches, chapters 1 and 9) and models of perceptual processes in music and language (such as Lerdahl and Jackendoff 1987, see also, for example, Hauser et al. 2002), and is reflected and incorporated in the rule base and algorithm (chapter 15). Furthermore, as discussed above, at each stage, there are often different relationships between components. The features and phrase parts were found to have different strengths affecting the resulting phrase-type. The different phrase types and phrase combinations therefore also have different strengths and consequently relationships between them.

It seems, therefore, that though there may be problems in applying a very rigorous definition of hierarchy to phrasing, the concept does help summarise the relationships between the different aspects discussed in this study.

13.8. Phrase length

There are many authors who approach phrases from the feature of an underlying regularity in their length. Moreover, length seems to be almost part of the definition of the phrase rather than a contributing feature. Here, phrase length is considered as one of the features and is therefore discussed in more detail in chapter 12.

The internal punctuation points, memory and phrase length

The internal 'punctuation points' discussed above, besides playing a role in structuring the phrase, may help in remembering the information. The capacity of working memory for processing music may be very limited, so the more opportunities the listener has to chunk information mid-way through a phrase, the longer (and more interesting?) a phrase can be. The perceived function of these 'internal punctuation points' can change as more information is given. Moreover, different types of internal punctuation can have different functions: close followed by extension (i.e. prolongation after close), thwarted close (i.e. delay of (prolongation before) close), mid-way point (e.g. antecedent consequent) depending on the nearest strong closes and the contributing features.

Effect of proximity between phrase starts

One of the 'structural' limitations on phrase length seems to relate to the proximity between strong or weak features. The definitions we now have of phrases, the analyses and the experimental results suggest that there seems to be an effect of proximity with regard to the strong phrase starts and phrase ends but less of an effect for the intermediary punctuation positions. Putting things another way, the intermediary punctuation positions can follow each other rapidly but they are likely to be relatively weak if near strong phrase ends or starts.

The distance from the previous phrase boundary seems to influence the importance of the features, especially, the distance from the nearest strong phrase boundary. The nearer we are to the previous strong, complete phrase end and phrase beginning, the less likely we are to have a similar phrase boundary. However, if we pass an unfulfilled phrase 'boundary' we may expect its true fulfilment later.

13.9 Summary

This study suggests that the phrase may be seen as a 'unit' of musical information, the main components of which are its the beginning, expectation of the end and end (and by implication the beginning of a new phrase or end of the piece) but that not all of these have to be clearly presented in each phrase. Other components may also be present including: beginning of the end and arrival at the resolution. The expectation is of two types, each being important in different circumstances. One type is driven by "template" features from musical experience regarding: length, symmetry and bar position and that they should contain a beginning and end. The other type is "internally" generated from musical features, including change or repetition signalling starts, harmonic and voice-leading, as well as changes in texture, igniting expectations within a phrase, their resolutions indicating phrase ends, and boundary-features of long notes and pitch jumps indicating phrase boundaries. These features and the expectations generated from them may or may not coincide with a "standard" unit. Thus the various phrase parts and, therefore, their indicative features and feature combinations have different functions.

The discussion has elaborated upon the internal structure of the phrase, normally consisting of phrase parts. These are dependent on the musical features and their combinations. The interplay of these aspects results in a large variety of phrase structures.

The discussion has also emphasised that the phrase is distinguishable from, but related to, other phrases in different ways: neighbouring phrases can be separated by gap features or may partially overlap in elided phrases; phrases may be related to each other in a sequential manner or in that of antecedent-consequent pairs; and more distant phrases may be related to one another through, for example, repetition.

These discussions lead to the suggestion that, phrases have phrase parts and belong to different phrase types. Different phrase parts have their own features, feature combinations and functions in different musical contexts. On this basis, the following chapter suggests a rule base for phrase part and phrase type identification.

Chapter 14

Features, Feature-combinations and Phrase Labels: *Elaboration*

14.1 A Rule Base

14.1.1 Introduction

The following rule base summarises the main relationships between features and phrase parts identified in the preceding chapters. These relationships were derived from the analysis of participants' responses to the different types of renditions of the case-study pieces, from musical analyses of the pieces and of performance contours and from theoretical approaches and observations in the fields of music analysis and psychology. Participants identified 'phrase-parts' (phrase start, end, beginning of the expectation of the end). Analyses of the responses lead to the identification of additional phrase parts (beginning of the end, arrival at the resolution, prolongation of end, end of end and phrase gap chapters 11 and 12) and 'phrase-types' (strong/weak), all systematically associated with musical features (chapters 3-6, 11 and 12).

The feature combinations (of different strengths, section 14.1.2) that appear with phrase parts and phrase types (section 14.1.3), therefore, form the rules of the following rule base. Some of these rules are also already explicit or implicit in other formal approaches (such as some of those in Lerdahl and Jackendoff, 1987, and others explored in chapter 8) or in music theoretical approaches (such as those of Koch, 1787, 1983 and Rothstein, 1989) but have not been combined previously. Their combination, however, is supported by the studies carried out (chapters 3-6, 11 and 12). The 'combined approach' allows the amalgamation of specific characteristics identified for the case-study pieces, and responses to them, with other observations, both specific to particular pieces and general theoretical ones discussed in the literature.

The rule base consists of states corresponding to different phrase parts (discussed in section 13.5). To move between states, certain "conditions" have to be met; the conditions being the presence of features or feature combinations (section 14.1.5). The states and the formalisation of the possibilities of moving between them enable specified possibilities of orders of rule application and several concurrent interpretations of the same input. As a whole, the rule base can be applied to a number of different musical textures (section 14.1.4). The presentation of the tables is discussed in sections 14.1.5-6 Chapter 15 shows the applications of these rules to new 'test' pieces.

14.1.2 Distribution and Strength of Features and Rules (Primary, Complementary and Online)

The observations of the distribution of different features (chapters 11 and 12) are formalised in the rule base. The different feature types (instantaneous, predictive and retrospective) are distributed among the different phrase parts. Predictive are usually in states $S_{2\text{-}5}$, instantaneous S_6, and retrospective, S_1.

The results of the current study (chapters 1-13) indicate that there are different types of features coinciding with different phrase parts and types leading to different types of rules. In chapter 11, the features were primarily analysed from the perspective of frequency, leading to two categories, A and B, with the second being divided into two further subtypes. Here, these categories are recast in terms of relative strength of the different features. Furthermore, the distinction between the two different types of Category B features is formalised.

There are a number of features that, regardless of the combinations they occur with, clearly signal a particular phrase-part such as a phrase start, a phrase end or an expectation for a phrase end (Category A features, chapter 11). These are: "exact repeat", "following explicit voice-leading" and "following resolved cadential progression". There are also feature-pairs that, regardless of the other features that they occur with, clearly signal a phrase part. These include any pair of: "Following cadential progression", "After explicit voice-leading", "Four bar template", "Following implicit voice-leading", "Following increase in rate of harmonic rhythm", or "Four bar template (from start of melody)". These features are here referred to as primary features and are part of primary rules (indicated by P in table 14.3, column 6 below).

Other features often occur with the primary features, complement them and strengthen the signal (mainly Category B features, chapter 11). These features are therefore referred to in the following as complementary features (indicated by the C in table 14.3 column 6 below). When occurring on their own, these complementary features indicate phrase-parts that allude to, or 'mimic', the clearer phrase-parts but usually do not provide enough evidence for a definite phrase start, end or expectation. In such cases they may contribute to a 'temporary' or secondary phrase part within a larger phrase (chapter 11.3.4). Several of these complementary features, such as change in motive, occur in different contexts (phrase start or expectation of the end etc.) and this is one of the reasons why, on their own, these features are not clearly indicative of particular phrase parts. The studies of chapters 1-11 have shown that these complementary features can contribute to confusion, but it seems that, at least in retrospect, their rôle is often clearer in certain contexts, depending on presence or absence of other features and the musical experience of listener. For example, some complementary features (such as pitch jump) occur so often throughout the pieces that they only possibly enhance other features but are not clearly indicative on their own. When on their own and far from any primary features however, they become more important and in these situations do indicate phrase parts particularly if they are located when a phrase part may be even weakly 'expected'. This study has shown that listeners also seem to have a number of expectations that function as 'checks' alongside the features. For example, if a phrase or phrase-part is 'getting too long', i.e. if no clear features have been detected for a long time, there is an increasing expectation for a new phrase part. In such cases, the complementary (usually weaker) features increase in significance if they coincide with the 'expected' positions or lengths. Such expectations are discussed below (table 14.2).

Furthermore, some complementary features and their respective phrase-parts seem to recede in importance in the listeners' memory when a section or the whole piece is finished. This occurs primarily in two situations: 1) In retrospect, the expectations are no longer remembered as strongly as the phrase starts or ends if expectations are generated and not fulfilled, and 2) If there are indications of phrase starts or ends by weak features, these may later be forgotten especially if there are primary features and therefore stronger phrase starts or ends near by. These are also Category B features: change in texture and motive, and short rests, chapter 11).

The rule base reflects these changes in importance in two ways, allowing a choice between different levels of detail in the interpretation: 1) Expectation features and rules are included with the other features and are applied along with all the others in the process of 'listening'. Once the analysis is complete, it is possible to see which expectations were fulfilled and which were not, and the latter may be excluded from a 'retrospective' analysis. 2) There is a series of possible phrase-start features that if encountered alone or with weak features may act as temporary 'punctuation positions' during listening, but which may be omitted in retrospect. These are: "Inexact repeat", "Change in texture", "Change in motive" and "Chord", are referred to as Online features (O), and are indicated by O in Table 14.3, column 6 below. During the piece, online (O) features are equivalent to primary (P) features while in retrospect they are equivalent to complementary (C) features.

Conversely, some features that in general were found not to belong to Category A, such as rests and change in motive, gain much more importance in certain pieces (characteristic of certain genres or instrumentation, such as the importance of rests in vocal music). In these cases, such features take on the certainty usually associated with Category A features.

The above discussion described three general 'categories' of rules (P, C and O). Within the P category there is also a gradation of rules. These gradations refer to the same features types as if on different 'settings', some are stronger and some are weaker. These usually concern harmonic, voice-leading or metrical information. For example, a cadence that is clearly resolved on to the tonic of the piece both in the bass and in the top part, is given more weight than a cadence that ends with non-tonic notes in the base or the voice part. This is taken into consideration particularly when 'connecting' together the stronger (usually longer) phrase parts or in retrospective analysis; the stronger phrase parts arrived at from these more weighty features, may stay in the retrospective analysis more than the weaker ones. Currently, the different 'settings' are presented as different rules for clarity. In an implementation they could be combined. Despite the different weightings assigned to different rules, this is an additive model.

On the basis of the results of chapter 11, it is possible to assign weightings to particular features. For example, "exact repeats" indicate phrase starts for 100% of occurrences or "following cadential progression" and "following increase in rate of harmonic rhythm" indicate expectations of phrase ends do so for 100%, or "Following cadential progression" indicate expectations of phrase ends do so for 92%. This temptation is resisted for several reasons including: though these features are named individually, one of the assumptions of this rule base is that any number of different features with the same scope and function may be interchangeable (see chapter 12) and a small group of case-study pieces have been used in gathering this numeric data.

In summary, there are three ways in which the outcomes of the rules are weighted:
1. Different rules have different relative weighting Primary (P), Complementary (C) and Online (O), and this weighting may change depending on whether the information is currently being heard, or whether it is seen only in retrospect.
2. P rules can be 'weaker' or 'stronger' usually depending on the relative strength of resolution
3. In general, the greater the number of strongly weighted rules on a given note, the stronger the note as that phrase-part.

If the rules support each other, then the different labels are kept as one interpretation. However, if there is conflict (e.g. if a phrase start falls between a beginning of the end and a prolongation) there are a number of possibilities:
1. If the conflict is between P and C or P and O rules, P has greater weight than C. During the Parse, O is equivalent in weight to P. In retrospect, O is equivalent to C. It is possible to add the setting that non-musicians often favour 'online rules' even in retrospect.
2. If the conflict is between different levels of P rules (PS = Primary, strong, PW = Primary, weak), the strong outweigh the weak.

3. If the conflict is between rules from the same category and strength, then those that are successful for notes that are closest to the notes identified by the running counters have greater weight.

Otherwise, all the labels should be retained as this conflict would indicate ambiguity.

For each rule, a number of different states can be reached. If rules leading to different states are successful, these are all kept and continue as separate parses. In the phrase label loop, all the successful states are considered together and compared, enabling ambiguities, and overlapping and elided phrases.

14.1.3 Phrase Parts, Labels and States

The different phrase parts are those that were explicitly used in the listeners' and annotators' studies and additional phrase parts that arose from the analysis of these results, the performance contours and the pieces (discussed in section 13.5). These additional phrase parts separate out the different components of the preparation for or expectation of the phrase end, the beginning of the end, the arrival at the resolution, the prolongation of the end, the end of the end and the phrase gap. As discussed in chapter 13, section 13.4.d, the beginning of the expectation of the end of the phrase and the beginning of the expectation of the end of the phrase often overlap. Therefore, these two are represented in the same states in the rule base (S_2).

The results of the current study indicate that not all of these phrase parts are necessary for phrases to be identified. However, if they do occur, they usually appear in certain combinations and sequences. The different relationships between the phrase parts are formalised in this rule base as 'states' and the relationships between them and the semantics of the different states is given in table 14.1 below.

14.1.4 Musical Textures

These rules are designed to be applicable to music of different textures: monophonic, music that has more than one part but in which all voices can be treated together, melody and accompaniment and more complex polyphonic music. Before starting the process of phrase analysis it is necessary to decide whether the piece should be treated as having one or more voices on the basis of the whole piece (Table 14.2). This can be done manually. The automation of this step is more complex. The method for identifying the different voices could begin with such methods as those described by Bregman (1990). These methods would have to be developed further in order to assess the level of independence between the voices and therefore the need to treat them separately or together. Initially, if there is more than one voice, the rules should be applied separately to each of them. The more the phrase parts identified for the different voices coincide, the clearer they are to the listener. In most pieces with solo and accompaniment texture, the phrasing of the solo is 'stronger' than that of the accompaniment: the phrases of the accompaniment can often be seen as anticipating or prolonging the solo phrases. However, there are more complex cases in which the relationship is not so simple (such as in the Bach Cantata 140 no. 4 below). If there is more than one voice and if it is not clear whether or not the different parts should be treated individually, the analysis can be carried out on the individual voices. If the features and phrase-part positions of the different voices coincide, they can be treated as one voice, if they do not, they should be treated as separate voices.

14.1.5 The rule base and the tables

The general formulation of the rule base is in the form of a finite state machine. The input to the machine is a set of features that are extracted from the music independently of the machine itself. The conditions for each rule are specified in terms of features that are assumed to be available and all the rules that may apply to a given state are checked sequentially. The output of each rule includes: $label, $location, $strength and state transferred to. The strengths of all of the applicable rules at a given state are compared and only the rule with maximal strength is applied. The $label and $location are included in the semantics of the states and are therefore somewhat redundant in terms of running the rule base. They are presented here for clarity and interpretation.

Table 14.3 shows the individual features and the phrase-labels that are associated with them. The features are grouped according to the phrase-parts that they indicate. The phrase-parts are presented according to the order in which they may normally be encountered in a piece but, as will be seen from the different state combinations discussed below, the order is rather flexible. If none of the expected orders materialise, it is suggested that listeners would be confused. This would be represented by the contradictory labels that would result from the rule base (see chapter 13 for discussion of identification of phrase parts and their order).

Combinations of different features for the most part complement one another – the more features there are that indicate the same phrase-part, the stronger the signal for that phrase part. For many of the features there may be several manifestations of the same general feature. For example, the different metrical location of voice-leading descent or cadential progressions, or different voicings of the final chord (tonic in the bass and the top part vs. non-tonic in both of these parts) can result in phrase parts of different strength. The distinction between the clear primary and complementary features, and the different manifestations of the same feature that would not co-occur mean that usually there should not be misleading contradictions between the rules below. Misleading here refers to a case in which an ambiguity would be suggested by the rules that does not exist for any listener.

Table 14.4 shows the performance features and the phrase-labels that are associated with them. These could be included in the same table as the musical features (table 14.3). However, here they are kept separate in order to highlight the conclusion from this study that they are not essential for phrase perception. The responses to the performances indicated that underlying tempo also plays a role in the 'lengths' of phrases identified: the faster the underlying tempo, the 'more notes' are included in each phrase. This is summarised in table 14.5.

A presentation of the most 'complete' sequence of phrase labels (derived the results gathered in previous chapters) follows. These phrase labels are then used to show what different phrase-part relationships may indicate and can be labelled with these indications. Table 14.6 shows this through a group of if/then rules. The first rule is that if the complete sequence of phrase-parts is identified, together they can be labelled a complete phrase. The following rules give more possible phrase-part combinations, both straightforward relations such as if a phrase start follows phrase end, this is labelled a confirmed phrase boundary, and more complex combinations such as conflicting phrase parts leading to contradictions and ambiguity. In more detail, if there is a contradiction between phrase parts that have been identified, it may be that all the phrase labels are equally strong (that they resulted from primary rules) in which case the combination is ambiguous. However, if there is a combination of both primary and complementary rules, the primary rules override the complementary ones – this would be decided in combination with the earlier tables (chapter 13).

This rule base is intended to represent aspects of the listeners' perception of phrase structure. Its aim is not to give one definitive 'answer' nor to resolve ambiguities but to identify them in cases in which a listener or performer might do so. For example, areas which are ambiguous for the listener may be identified if the application of these rules results in combinations of musical features and their corresponding phrase parts that contradict each other or are not accounted for in the phrase labels and combinations as presented, for example in Table 14.6. As the experimental results have shown, there can be several possible phrase-structure interpretations for each piece. There are a small number of key phrase positions that recur in most interpretations. There are additional phrase positions that are noted less often and by different proportions of listeners and are therefore here regarded as weaker. The following rule base distinguishes between the strong and weak positions and their respective features but does not automatically provide a systematic relationship between features and weaker positions. Moreover, as discussed above, even the same listener may hear and respond with different phrase interpretations each time the piece is heard (even when listening to the same MIDI rendition). For these reasons, the different phrase structure interpretations are not ranked.

All of the musical features have different characteristics in the way they are perceived individually and in their rôle in the perception of phrase structure: they can be expected, instantaneously heard or identified retrospectively, and their implications for the phrase perception can be predictive, instantaneous and retrospective (chapters 12-13). These implications are given at the start of each phrase part in tables 14.3 and 14.4. These different characteristics combine and affect the way the phrase is heard. The different temporal effects influence their individual perception, the way their sequence, and therefore, the phrase are perceived.

Phrases tend to be organised in the relations of table 14.6 or there maybe 'shorter' phrases over the same areas as the longer ones. These could be referred to as 'subphrases'. However, the results of the previous chapters showed that for some participants and analysts these seemed to be considered the main phrases. Furthermore, participants, analysts and performers tend to emphasise both long and short phrases, only sometimes systematically identifying sub-phrases within longer ones.

For example, if $current_state = S_0 or S_1 or S_2 or S_3, then the following rule applies (see table 14.3, Arrival at the resolution section, first rule). The rule is successful if there is an arrival on the resolution of a cadence (column 2). If there is no resolution, the next rule is attempted. If there is a resolution, this indicates that this may be the phrase end (column 3), which is located at the end of the resolution of the cadence (column 4). As discussed above, resolved cadences are primary features, so the relative strength is P. The states transferred to is S_3 (arrival at resolution). From this rule then, current states moves to (or returns to) states 3, 4, 5, 6 or 1 (see table 14.1 below). In the Bach Passion for example, we arrive on the first note of bar 13 in state 2, expecting the end of a phrase and indeed the tonic arrives on the first beat of the bar.

Conditions		**Outputs**			
1	2	3	4	5	6
Current State	**Features**	**Label**	**Location**	**Relative strength**	**State Transferred to**
S_0 S_1 S_2 S_3	Arrival on resolution of cadence	Resolution + ? Phrase end	Phrase end at end of resolution of cadence	P	$\boldsymbol{S_3}$

This rule, like all the others in the rule base, is derived from a number of different sources. For example, in terms of the case-study pieces, almost all have at least one position where a phrase end response or implication in the performance contours coincides with a cadence (such as the Bach Passion, bars 12-13, the Mozart Sonata bars 7-8 or Mozart Aria, bars 8-9). In terms of music theory, Koch and Rothstein both emphasise the importance of a resolved cadence (chapter 9).

14.1.6 Presentation of the rules in tables 14.1, 14.3 and 14.4

As mentioned above, the first two tables separate the 'musical features' (those available from MIDI, **Table 14.3**) from performance features (**Table 14.4**). Within each table, each rule is an '**if/then**' rule: **if** the listed **features (Tables 14.3 and 14.4 Column 2)** are identified, **then** the given **label (Tables 14.3 and 14.4 Column 3)** can be placed at the **location** indicated **(Tables 14.3 and 14.4 Column 4)**. The rules are grouped according to the phrase-parts to which they apply. Similar or even the same features can appear more than once. Each rule is also placed in the context of states. States and the relationships between them summarise the different possible relationships between phrase parts; given a particular phrase part, which phrase parts can be expected to follow (**Table 14.1**). As can be seen from the table, three of the states can be followed by a return to the same state. It is possible that there may be repeated signals for the beginning of the expectation of the end of the phrase/beginning of the end of the phrase (S_2), the arrival at resolution (S_3) and particularly the prolongation of end (S_4). This rule base is intended as a parser rather than a generator. For a generating rule base, there would be an additional rule to enable a timely exit from such loops. As a parser, it would be expected that this loop would be exited automatically moving on to the next state and at worst case, musical pieces should not continue indefinitely. In Tables 14.3 and 14.4 the states indicate at which parts of the phrase each rules are applicable: where the listener is "now" in the phrase in order for each rule to be applied (**Tables 14.3 and 14.4, Current State**, **Column 1),** and where in the phrase that rule would lead to **(Tables 14.3 and 14.4, State Transferred to, Column 5).**

In summary, each rule is presented with its own conditions and actions: conditions are the $states and $features, and actions are the $label, $location and $state transition. As mentioned above, the tables also show whether the features and therefore rules are primary or complementary and whether the features are applicable only in the online phase of phrase perception (**Tables 14.3 and 14.4 Column 6 - P, C, or O)**.

As discussed above, table 14.1 presents the different possible combinations of states and which phrase parts they indicate.

Table 14.1 State Transitions table

From	To
S_0	S_1 S_2 S_3 S_5 S_6
S_1	S_2 S_3 S_5
S_2	S_2 S_3 S_5
S_3	S_3 S_4 S_5 S_6 S_1
S_4	S_4 S_5
S_5	S_6 S_1
S_6	S_1

Key for state transitions table

S_0 – initial state
S_1 – phrase start
S_2 – beginning of the end of the phrase (with beginning of the expectation of the end)
S_3 – arrival at resolution
S_4 – prolongation of end
S_5 – end of end
S_6 – phrase gap

Table 14.2 shows different possible textures and how they should be treated in this rule base. In addition to the musical features, this study has shown that listeners also have a number of expectations that function as 'checks' alongside and in addition to the features. For example, if a phrase or phrase-part is 'getting too long', i.e. if no clear features have been detected for a long time, the weaker (complementary) features increase in significance if they coincide with the 'expected' positions or lengths. These are also summarised in table 14.2

Table 14.2 What is the general texture of the piece? Underlying assumptions about phrase length (Checks)

Texture - Does the piece have one or more voices and what is the relationship between the voices:

If:	***Then:***
One voice	treat as monophonic, all rules applicable to this voice
Solo and accompaniment	treat as two voices, all rules applicable to solo and accompaniment separately
Multi voice but together	treat as homophonic, all rules applicable to all as one voice
Polyphonic	all rules applicable to each voice separately

The system can be entered anywhere. Only if it is the beginning of the piece is it *a priori*, S_0

Checks - In order to capture the following basic expectations, a 'clock' is started from the beginning of the piece and from the beginning of each phrase start, and expectation for a phrase boundary increases after appropriate real time or number of bars pass.

Phrase length	The experimental results indicate that phrase lengths seem 'reasonable' up to 11 seconds though most are shorter. According to some we have an internal limit on the amount of information in each 'unit' or, in this case, phrase (chapter 1).
Template	Both theorists (e.g. chapters 1, 8 and 9) and some listeners in their written responses indicate that there is a basic weak expectation of even-bar-length phrases, usually 4, 8 or 2 bars long. However, both theorists and listeners show that this is the 'basic' length that can be modified, usually extended but sometimes shortened, primarily through elision.
Symmetry	There seems to be a stronger expectation of phrases to be of similar lengths (or divisions or multiples of that length), again usually measured in bars. Moreover, there also seems to be an expectation for successive phrase starts to occur in the same part of the bar or near the same strong beats.

Table 14.3 is the main table showing the states (column 1), conditions for their applications (features of column 2), and the outcomes; possible movements between them (column 6) and the different labels indicating phrase part, its location and its relative strength (columns 3-5).

The first rule applies to piece beginnings or the first beginning of the melody. The rest of the phrase start rules (S_1) are given at the end of the rules, reflecting the order of phrase ends followed by starts. If the analysis starts partway through a piece, states are concluded using the features. At the end of the table, a list of underlying checks is given.

Table 14.3, Musical Features and phrase labels

Conditions		Outputs			
1	2	3	4	5	6
Current State	**Features**	**Label**	**Location**	**Relative strength**	**State Transferred to**
S_0	***First phrase start (S1)***				
	At start of piece,				
	beginning of piece	Phrase start	Phrase start at start of piece	P	***S_1***
	beginning of solo line (if later)	Phrase start	Phrase start at start of melody?	P	***S_1***
Beginning of the end of the phrase (S2)					
All except last two: Retrospective identification, Predictive of phrase end					
S_0 S_1 S_2	Beginning of cadential progression	Expectation of end begins	From beginning of cadence	P	***S_2***
	Strength of Cadential progression				
S_0 S_1 S_2	Beginning of cadential progression	Strong Expectation of end on strong beat of bar	“”	PS	***S_2***
S_0 S_1 S_2	Beginning of cadential progression	Weak Expectation of end on weak beat of bar	“”	PW	***S_2***
S_0 S_1 S_2	Beginning of voice-leading descent	Expectation of end begins	From beginning of voice-leading descent	P	***S_2***

$S_0 S_1 S_2$	Beginning of voice-leading descent	Strong Expectation of end on strong beat of bar	“”	P	$\boldsymbol{S_2}$
$S_0 S_1 S_2$	Beginning of voice-leading descent	Weak Expectation of end on weak beat of bar	“”	P	$\boldsymbol{S_2}$
$S_1 S_2$	Change in texture Fuller texture, associated with cadential progression	Expect end of Phrase begins	From change in texture	O	$\boldsymbol{S_2}$
$S_1 S_2$	Change in motive	Expect end of Phrase begins	From change in motive	O	$\boldsymbol{S_2}$
$S_0 S_1 S_2$	Use of ‘standard’ ending motive	Expect end of Phrase begins	From beginning of motive	P	$\boldsymbol{S_2}$
$S_1 S_2$	Repetition of ending motive used previously in piece - associated with phrase ends primarily a particular way of cadencing	Expect end of Phrase begins	From beginning of motive	P	$\boldsymbol{S_2}$
$S_0 S_1 S_2$	Increase in rate of harmonic rhythm associated with cadential progression	Expect end of phrase begins	From change in tempo	C	$\boldsymbol{S_2}$

Arrival at resolution (S3)

Here '? Phrase end' indicates that the phrase end is only a possibility (potential phrase end): if it is followed by phrase 'gap' or phrase start, this position is retrospectively confirmed as the end, if it is followed by prolongation or end of the end, it is not.

All except last two: potentially expected, retrospective

$S_0 S_1 S_2 S_3$	Arrival on cadence resolution	Resolution + ?Phrase end	Phrase end at end of resolution of cadence	P	$\boldsymbol{S_3}$
$S_0 S_1 S_2 S_3$	Arrival on resolution of voice-leading	Resolution +?Phrase end	Phrase end at end of resolution of voice-leading	P	$\boldsymbol{S_3}$
$S_0 S_1 S_2 S_3$	Arrival at end of theme	Resolution +?Phrase end	Phrase end at end of theme	P	$\boldsymbol{S_3}$

Strength of progression

$S_0 S_1 S_2 S_3$	Arrival on cadence resolution, voice-leading or end of theme on strong part of bar	Strong Resolution +? Phrase end	Phrase end at resolution	P	$\boldsymbol{S_3}$
$S_0 S_1 S_2 S_3$	Arrival on cadence resolution, voice-leading or end of theme on weak part of bar	Weak Resolution +? Phrase end	Phrase end at resolution	P	$\boldsymbol{S_3}$

Prolongation of end (S4)

All: retrospective:

$S_3 S_4$	Resolved cadential progression followed by same note/chord	Prolongation	Prolongation on repeated note/chord	P	$\boldsymbol{S_4}$

S_3 S_4	Penultimate note lengthened (suspensions)	Prolongation	Prolongation on lengthened note	P	***S_4***
S_3 S_4	Repeat of phrase ending	Prolongation	Prolongation at repeat	P	***S_4***
S_3 S_4	Extension prolonging onto strong beat	Strengthening phrase end	Prolongation during extension to strong beat	P	***S_4***

End of end (S5)
In many ways the rules here are interchangeable with those in S_3. However, as some of these could only arrive after S4 they are grouped separately
All potentially expected

S_4	Prolongation end after resolution	Phrase end	at end of resolution	P	***S_5***
S_4	End of resolution of cadence	Phrase end	at end of cadence	P	***S_5***
S_0 S_1 S_2 S_3 S_4	Arrival on resolution of cadence	Phrase end	end of phrase at resolution	P	***S_5***
	Strength of progression				
S_0 S_1 S_2 S_3 S_4	arrival on resolution of cadence and root in melody and bass	Phrase end	Strong phrase end	PS	***S_5***
S_0 S_1 S_2 S_3 S_4	arrival on resolution of cadence and root only in bass and delayed or no resolution in melody	Weaker Phrase end	Weak/Temporary phrase end	PW	***S_5***
S_0 S_1	arrival on resolution of cadence:				

S_2 S_3 S_4 metrically strong close	Phrase end	stronger	P	$\boldsymbol{S_5}$
S_0 S_1 arrival on resolution of cadence: S_2 S_3 S_4 metrically weak close	Weaker Phrase end	possibly weaker	P	$\boldsymbol{S_5}$
S_0 S_1 arrival on resolution of explicit S_2 S_3 S_4 voice-leading	Phrase end	Phrase end at resolution	P	$\boldsymbol{S_5}$

Phrase 'Gap' (S6)
Instantaneous:

S_0 S_3 S_5 Long rest	Phrase boundary Distinct phrases	Phrase boundary during rest new start on note after rest	P	$\boldsymbol{S_6}$
S_0 S_3 S_5 Short rest	Phrase boundary Less separate phrases	Phrase boundary during rest new start on note after rest	O	$\boldsymbol{S_6}$
S_0 S_3 S_5 Pitch Jump	Phrase boundary	Phrase boundary between notes new start on second note of pitch jump	C	$\boldsymbol{S_6}$
S_0 S_3 S_5 Resolved cadential progression followed by rest (and/or pitch jump)	Prepared Phrase end, rest, phrase start	Phrase end, end of cadential progression new phrase after rest	P	$\boldsymbol{S_6}$
S_0 S_3 S_5 Distant from harmonic resolution but rest (and/or pitch jump)	Surprise Phrase end, rest,	Phrase start phrase after rest Phrase boundary and breath	P	$\boldsymbol{S_6}$

Phrase start (S1)

Instantaneous and retrospectively identified:

S_0 S_3 S_5 S_6 Exact repeat	Phrase start	Phrase start at start of repeat	PS	***S_1***
S_0 S_3 S_5 S_6 Repeat of beginning	Phrase start Strong, immediately Clear. phrase start	Phrase start at start of repeat	PS	***S_1***
S_0 S_3 S_5 S_6 Repeat of non-beginning	Weaker Phrase start	Phrase start at start of repeat	C	***S_1***
S_0 S_3 S_5 S_6 Change in texture/motive	Phrase start	new phrase at change	O	***S_1***
Strength of change in texture				
S_0 S_3 S_5 S_6 Big (long-term) change in texture	Phrase boundary Distinct phrases	Phrase start at beginning of change	OS	***S_1***
S_0 S_3 S_5 S_6 Small (short-term) change in texture	Phrase boundary beginning of change	Phrase start at Less separate phrases	OW	***S_1***
S_0 S_3 S_5 S_6 Return to original rhythm/tempo	Phrase start	Phrase start at change in rhythm/tempo	P	***S_1***
S_0 S_3 S_5 S_6 Inexact repeat follows resolved cadential progression	Phrase start beginning of repeat	New phrase at	O	***S_1***
S_0 S_3 S_5 S_6 Text: new sentence/clause	Phrase start	New phrase at start of text	P	***S_1***

Table 14.4 presents the same structures, this time on the basis of the different performance features

Table 14.4, Musical features and phrase labels from 'Performances'

Performance Features

Performance features both represent phrase interpretations and, as was seen in the listeners' study, direct listeners' perceptions, highlighting or playing down phrase structures identified in the MIDI experiment. This is done through:

Beginning of the end of the phrase (S1)					
S_0 S_1	Ritardando	Expect end of phrase	From beginning of Ritardando	P	$\mathbf{S_2}$
S_0 S_1	Diminuendo	Expect end of phrase	From beginning of Diminuendo	P	$\mathbf{S_2}$
Phrase 'Gap' (S6):					
S_0 S_3 S_5	Lengthened note	Phrase boundary	Long note is last note of phrase or first note of new phrase	C	$\mathbf{S_6}$
S_0 S_3 S_5	Breath	Phrase boundary	Breath is phrase boundary	P	$\mathbf{S_6}$
Phrase start *(S1)*:					
S_0 S_3 S_5 S_6	Long term change in tempo or dynamics	Phrase start	Phrase start at start of change	P	$\mathbf{S_1}$

Table 14.5 summarises the dependency between tempo and phrases

Table 14.5

Underlying Tempo

In general, the underlying tempo of the performance seems to influence the lengths of phrases (chapters 3 and 10). In general the following relationships hold:

Faster tempo	longer phrases (i.e. more notes included in same phrase), less phrases in the piece
Slower tempo	shorter phrases (i.e. less notes included in same phrase), more phrases in the piece

The phrase and its parts

The nature of the phrase parts and the relation between them can be complex. There can be several different combinations of phrase starts; if presented with a phrase part, there are several different options for continuation. The most complete and elaborate sequence would be the following one, but it is rarely fully represented.

Phrase start - Expectation of phrase end – Beginning of the end – Prolongation – Arrival at Resolution – Prolongation – End of the end – Breath – Phrase start.

The following table shows rules that can were derived from the observed sequences in the test pieces.

Table 14.6

Phrase Label Combinations (on basis of features/phrase labels):

If	***Then***
All phrase parts follow in sequence (above)(S_1-S_6)	Complete phrase
Phrase start (S_1) follows phrase end (S_3, S_5 or S_6)	Confirmed phrase boundary
Phrase end (S_3, S_5 or S_6) not followed by clear phrase start (not S1, or for S_3 not followed by S_4).	Contradictory, weakened phrase end
Phrase end expected and achieved (combination of S_3, S_4, S_5, and S_6)	Fulfilled expectation
Phrase end expected but not achieved (S_2 not followed by S_3, S_4, S_5 or S_6)	Unfulfilled expectation
Several repeats (exact/inexact)	Sequence
Phrase start (S_1) precedes phrase end (S_3 or S_5)	Overlapping phrases
Phrase start (S_1) coincides with phrase end (S_3 or S_5)	Elided phrase boundary
Conflicting phrase parts	There are contradictory features and therefore phrase parts – Ambiguous

14.2 An Algorithm

14.2.1 Introduction

The above rule base derived from the results of the preceding chapters is formalised as the following algorithm which will be applied to new 'test' pieces in chapter 15.

14.2.1.1 Overall Structure of the algorithm

The algorithm is made up of three parts: I. initialising, II. the main loop, and III. concluding and presenting the outcome at end of the input. The main loop consists of rules using features to identify phrase parts, followed by rules using phrase parts to identify phrase types.

14.2.1.2 Input and Output

The algorithm takes as input the name and length of each note in the extract or piece. It uses the musical features (identified manually or automatically) to annotate the phrase part label both in words ($label) and symbolically ($State). The State variable formalises the idea of the six different phrase parts described above (chapter 13). It is used, in the first part of the main loop, to move among the 'grammatically' relevant groups of rules from note to note and, in the last part of the main loop, to annotate the phrase labels ($phrase_labels). Several of the rules rely on more than one note and may be applicable to areas rather than specific notes.

Alongside the annotation of the $State and the $Label for each note that has a successful rule, the location $Location and category $Category are recorded. The $Category is used throughout the algorithm to assess the relative strength of rules and the resulting phrase parts and in reading the output. All these variables are used in reading the output reflecting different aspects of its conclusions.

14.2.1.3 Running the algorithm

In general, the algorithm is applied by going through the input note by note and attempting to apply each rule that is applicable given the state reached ($current_state) from the previous note. Even if there is success in one rule, all the following rules are applied. The number of successful rules for each position is one of the ways in which relative strength of each note as a possible phrase part is determined.

14.2.1.4 Presentation of main if-then rules

As the main loop consists only of if-then statements, using the same variables with different possible instantiations, the rules are presented in the form of a table. At the top of the table, the if-then condition is given, and at the top of each column the different variables are provided. As discussed above, for each of the notes the $label, $Location, $Category and $State are recorded.

14.2.1.5 Presentation of the rules in relation to their states

As many rules are applicable at many different possible states in a piece, the most concise presentation of the rules shows them in the order of the states that they can lead to. The states that they are applicable from are given at the start of each section of rules on the left in the form $current_state = $$S_n$. If a rule is successful, the current state changes to the State in the right-most column. When the rules are then applied to the next note, only those rules under the new State label are applied. The numbers before the point refer to the state that the different rules result in and are used for reference in the next chapter.

14.2.1.6 Adding the Phrase combination labels

The rules for adding phrase labels are, as in the main loop, presented in table form. If the State combinations listed in the *if* column are found, then the labels in the *then* column are added. Alternative options in case of conflict are given after the main table. As in the main loop, the information which results from these rules is sometimes complimentary, so all the rules should be applied, and more than one label may be successful.

14.2.1.7 Reading the 'output'

There are two outputs: the online response (reflecting the note-by-note assignments), and the retrospective one (reflecting the interpretations of the extract or piece as a whole). Within each of these interpretations, there will be some notes with stronger and others with weaker weightings for the different phrase parts. In general, 'long phrases' can be identified by connecting the more strongly weighted notes, and the 'short phrases' can be identified by also connecting all the weaker notes (section 14.1.2 above).

14.2.1.8 Notes on presentation

Comments specific to rules are given on the right following a #, variables are preceded by a $, and loops are open and close with { and } respectively.

The if/then rules are listed according to the groups of states they lead to ($State). At the top of each group the list of states that the rules apply for (i.e. the state that was reached with the previous successful rule $current_state) is given in red. By its nature, this is an exclusive or: for each rule only one state was reached. These are presented as lists of possible states to which the rules apply. For example: if $current_state is = either S_0 or S_1 or S_2.

On the other hand, within each list of rules, all the rules should be applied one after the other. Therefore, they are separated by inclusive Ors.

In the Phrase label loop, different combinations of phrase-parts ($States) are listed. In this loop the lists of states compiled by the previous loop are compared to these combinations and if there is a match, the sections are labelled with the relevant phrase label $phrase label. These are presented as lists of states. It is possible that the different states are returned to more than once in the lists compiled in the previous loop. For this Phrase label loop all of the repetitions are concatenated (as indicated by the Kleene stars$^+$). For example, the first list (S_0 S_1 S_2^+ S_3^+ S_4^+ S_5 S_6) means that all the states listed should appear and follow each other in that order and there can be 1 or more occurrences of states 2, 3 and 4. Further regular expressions are used for conciseness:

&& denotes 'and',
|| denotes 'or',
¬ denotes 'not' followed by, and
S_n^+ denotes one or more instances of S_n.

For example, in the previous example, all the states had to be present, in the following example any number of repetitions of states 3 and 4 should follow one another. These should be followed by either state 5 or 6 or both: S_3^+ S_4^+ S_5 || S_6. In the following list, state 2 is not followed by any of states 3 to 6: S_2 ¬ (S_3^+ || S_4^+ || S_5 || S_6).

14.2.2 The Algorithm

I. Initialising – no. of voices, where in the piece are we, setting the checks counters,
#There are three parts to the initialisation that can be done concurrently. One decides whether to treat the input as one voice, two-voice, homophonic, or polyphonic. #Another decides where in the piece the input begins: if it is the beginning of the piece, this is also assumed to be the beginning of a phrase, so the first phrase start can #be assumed to have happened, if it is not the beginning of the piece, the #feature rules should begin to run from the beginning until one is applicable. The third sets all #the checks counters at 0 or at the position in the bar in which the piece or section starts.

#A. How many voices are there?

If	***Then:***	
One voice	foreach $note, apply all rules	#$note is each note in the input:
Multi-voice but together	$note = $notes_onset foreach $notes_onset, apply all rules	#$treat each note onset separately
Solo and Accompaniment	Solo = list(notes_of_voice1), Accompaniment = list(notes_of_voice2), foreach $note in (notes_of_voice1) and in ($notes_of_voice2)	#$note notes_of_voice1, are as in #$note in one voice. $note in apply all #rules to for each position to each voice #in turn $notes_of_voice2 are as in #$notes_onset
Polyphonic	voice 1 = list(notes_of_voice1), voice 2 = list(notes_of_voice2), voice 3 = list(notes_of_voice3), voice *n* = list(notes_of_voice*n*), foreach $note in (notes_of_voice*n*), apply all rules	#treat each voice as a parallel run

#B. Where in the piece are we?

If Start of piece **then** $current_state = $\boldsymbol{S_1}$ #Because we know it's a phrase start
If non-start of piece **then** $current_state = $\boldsymbol{S_0}$ #Because we don't know where we are

#C. Set 'checks' counters.

1. Phrase length

```
If Start of piece then $current_state = S1          #Weak expectation of phrases not to last more than 11 sec (at extreme)
        Set $time_in_secs = 0
        Set $increase_weight_of_end/start = 0       #Currently there is no increased preference for a phrase end or
                                                    #start. If 11 seconds is reached, this is changed to 1 and the #relative
                                                    #weighting of phrase ends or starts is increased. These
                                                    #$'increase_weight_of_ variables are all binary; either increase #or do
                                                    not change the relative weighting.

Else Set $time_in_secs = [unknown]                  #We have to wait until the first phrase start is found
    Set $increase_weight_of_end/start = 0           #For this setting a way of counting real time would be necessary. If the
                                                    #input is MIDI, one way #would be to record the beats per minute and
                                                    #keep a #counter of how many beats and therefore #seconds have
                                                    past
```

2. Beat number

```
If $current_state = S1                              #Weak expectation of following phrases to begin on the same
                                                    #beat
        Set $beat_number_started = [no. of beat in the bar]     #So keep a record of the beat in the bar of the first phrase start
        Set $counting_beat_number = [no. of beat in the bar] #and a running record of the beat in the bar in the piece
        Set $increase_weight_of_start = 0
Else    Set $beat_number_started = [unknown]        #We have to wait until the first phrase start is found
        Set $counting_beat_number = [unknown]       #We count these in terms of bars and beats
        Set $increase_weight_of_start = 0
```

3. Bar number

```
If $current_state = S1
        Set $bar_number = 1                         #Weak expectation of phrases to be of even and equal length
        Set $length_in_bars = 0                     # keep a record of phrase length in bars
        Set $increase_weight_of_start = 0
Else    Set $bar_number = [unknown]                 #We have to wait until the first phrase start is found
        Set $increase_weight_of_start = 0
```

4. Interpretation log

If $current_state = ***$S_1$***

Print $position_label = Phrase start #If the start of the input is the start of the piece, we can assume that it is the

Else Set $position_label = [unknown] #start of the first phrase. Otherwise, we have to wait.

II. Running the input (one piece at a time): **#This is the main loop**

$no_of_rules_applied = 0 #foreach note re-initialise the no. of successful rules to 0

While (not end of input) {

#Resetting the counters

$no_of_rules_applied+=1

$counting_beat_number+=1

if($counting_beat_number = Integer){ #Bars begin on a new integer, so if a new integer is reached, the next bar has been reached and

($bar_number+=1) #this is recorded

}

Foreach $note:

if $current_state = either **S_0 or S_1 or S_2**

	And $features =	***Then* add:**			
Rule No.	**$Features**	**$Label**	**$Location**	**$Category**	**$State** transferred to
2.1	beginning of cadential progression (by the V if not preparatory V)	EOP begins	From beginning of cadence	P	***S_2***
	OR				

Strength of Cadential progression					
2.1.1	beginning of cadential progression on strong beat of bar	Strong EOP	“”	PS	***S₂***
	OR				
2.1.2	beginning of cadential progression on weak beat of bar	Weaker EOP	“”	PW	***S₂***
	OR				
2.2	beginning of voice-leading descent	EOP begins	From beginning of voice-leading	P	***S₂***
	OR				
Strength/type of progression					
2.2.1	beginning of voice-leading descent on strong beat of bar	Strong EOP	“”	P	***S₂***
	OR				
2.2.2	beginning of voice-leading descent on weak beat of bar	Weak EOP	“”	P	***S₂***
	OR				
2.2.3	voice-leading (not to tonic)	EOP begins	From beginning of voice-leading	W	***S₂***
	OR				
2.3	change in texture. Fuller texture, associated with cadential progression	EOP begins	from change in texture	O	***S₂***
	OR				

2.4	change in motive	EOP begins	from change in motive	O	***S_2***
	OR				
2.5	'standard' ending motive	EOP begins	from beginning of motive	P	***S_2***
	OR				
2.6	repetition of ending motive used previously in piece Motives become associated with PEs, primarily a cadence type	EOP begins	from beginning of motive	P	***S_2***
	OR				
2.7	increase in rate of harmonic rhythm associated with cadential progression	EOP begins	from change in tempo	C	***S_2***

if $current_state is = either **S_0 or S_1 or S_2 or S_3**
and $features is = to:

3.1	Arrival on resolution of cadence	Resolution + ? PE	PE at end of resolution of cadence	P	***S_3***
Strength of progression					
3.1.1	arrival on resolution of cadence root in melody and bass	Strong PE	""	PS	***S_5***
	OR				
3.1.2	arrival on resolution of cadence root in bass	Strong PE	""	PS	***S_5***
	OR				
3.1.3	arrival on resolution of cadence root only in bass and delayed or no resolution in melody	Weak/ Temporary PE	""	PW	***S_5***

	OR				
3.1.4	arrival on resolution of cadence, no root in bass	Weak/Temporary PE	""	PW	***S_5***
	OR				
3.2	Arrival on resolution of voice-leading	Resolution +? PE	Phrase end at end of resolution of voice-leading	P	***S_3***
	OR				
3.3	Arrival at end of theme	Resolution +? PE	Phrase end at end of theme	P	***S_3***
	OR				
Strength of progression					
3.1.1	Arrival at resolution of cadence, voice-leading or end of theme on strong part of bar	Strong Resolution +? PE	Phrase end at resolution	PS	***S_3***
	OR				
3.1.2	Arrival at resolution of cadence, voice-leading or end of theme on weak part of bar	PE on weak part of bar Weak Resolution +? Phrase end	Phrase end at resolution	PW	***S_3***

if $current_state is = either **S_3 or S_4**
and $features is = to:

4.1	resolved cadential progression followed by same note/chord	Prolongation	Prolongation on repeated note/chord	P	***S_4***
	OR				

4.2	Penultimate note lengthened or repeated (suspensions)	Prolongation	Prolongation on lengthened or repeated note	P	***S_4***
	OR				
4.3	Repeat of phrase ending	Prolongation	Prolongation at repeat	P	***S_4***
	OR				
4.4	Extension prolonging	Strengthening PE	Prolongation during extension to strong beat	P	***S_4***

if $current_state is = either **S_0 or S_1 or S_2 or S_3 or S_4**
and $features is = to:

#The rules here are, to some extent, interchangeable with those leading to S_3.
#However, as some of these could only arrive after S4 they are grouped separately

5.1	End of prolongation after resolution	PE	PE at end of resolution	P	***S_5***
	OR				
5.2	End of resolution of cadence and/or voice-leading	PE	PE at end of cadence	P	***S_5***
	OR				
5.3	arrival on resolution of cadence and/or voice-leading	PE	PE at resolution	P	***S_5***
	OR				
Strength of progression					
5.3.1	arrival on resolution of cadence root in melody and bass	PE, Strong PE	“”	PS	***S_5***
	OR				

5.3.2	arrival on resolution of cadence root only in bass and delayed or no resolution in melody	Weak/Temporary PE	“”	PW	***S_5***
	OR				
5.3.3	arrival on resolution of cadence, no root in bass	Weak/Temporary PE	“”	PW	***S_5***
	OR				
5.3.4	arrival on resolution of cadence:. metrically strong close	Stronger PE	“”	PS	***S_5***
	OR				
5.3.5	arrival on resolution of cadence: metrically weak close	Weaker PE	“”	PW	***S_5***
	OR				
5.4	arrival on resolution of explicit voice-leading	PE	“”	P	***S_5***

if $current_state is = either **S_0 or S_3 or S_5**
and $features is = to:

6.1	long rest	Phrase boundary (gap), Distinct phrases	Phrase boundary during rest (gap), PS on note after rest	P	***S_6***
	OR				
6.2	short rest	Phrase boundary	Phrase boundary during rest (gap) new start on note after	P	***S_6***

		(gap), Less separate phrases	rest		
	OR				
6.3	pitch jump	Phrase boundary (gap)	Phrase boundary between notes new start on second note of pitch jump	C	S_6
	OR				
6.4	resolved cadential progression followed by rest (and/or pitch jump)	Prepared PE, rest (gap), PS	PE, end of cadential progression PS after rest (gap)	P	S_6
	OR				
6.5	distant from harmonic resolution but rest (and/or pitch jump)	Surprise PE, rest (gap), Phrase boundary and breath	PS after rest (gap)	P	S_6

if $current_state is = **$S_1$**
and $features is = to:

6.6	long note	phrase boundary	PE on long note, PS, next note	W	S_6

if $current_state is = either **S_0 S_3 S_5 or S_6**
and $features is = to:

Foreach of the following rules**{**

If $state = S_1

Then

Set $time_in_secs = 0
Set $beat_number = [no. of beat in the bar]
Set $bar_number = 1
Print $position_label = Phrase start Weak expectation of phrases to be of even and equal length

1.1	repeat	PS	PS at beginning of repeat	P	S_1
Type of repeat					
1.1.1	exact repeat	PS	PS at beginning of repeat	P	S_1
	OR				
1.1.2	repeat of beginning of piece (inexact or exact)	Strong, immediately clear PS	PS at beginning of repeat	PS	S_1
	OR				
1.1.3	Repeat of non-beginning	Weaker PS	PS at beginning of repeat	C	S_1
	OR				
1.1.4	inexact repeat follows resolved cadential progression	PS	new phrase at beginning of repeat	O	S_1
	OR				
1.1.5	inexact repeat	PS	new phrase at beginning of repeat	C	S_1
	OR				
1.2	change in texture/motive	PS	new phrase at change in texture/motive	O	S_1
Strength of change in texture:					
1.2.1	Big (long-term) change in texture	Phrase boundary, Distinct phrases	PS at beginning of change	OS	S_1
	OR				

1.2.2	Small (short-term) change in texture	Phrase boundary, Less separate phrases	PS at beginning of change	OW	$\boldsymbol{S_1}$
	OR				
1.3	return to original rhythm/tempo	PS	PS at change in rhythm/tempo	P	$\boldsymbol{S_1}$
	OR				
1.4	text: new sentence/clause	PS	PS at start of text	P	$\boldsymbol{S_1}$

#Check the checks 'counting' variables:

If ($time_in_secs ≥ 11){ #From the data, it seems that 11 sec is an upper limit

$increase_weight_of_end/start = 1 #If this variable gets to it's limit and if there is success in the rules in the next loop in

#phrase ends or starts, the relative weights of these should be increased

Print "This phrase is getting too long"

}

If ($counting_beat_number = $beat_number_started) { #If the running beat number is the same as that of the #first phrase, there is a slight preference for a phrase start #so increase the weights of $\boldsymbol{S_1}$ rules.

$increase_weight_of_start = 1

Print "current position in the bar = the first phrase start

}

if ($bar_number = $length_in_bars) { #If the running $bar_number is the same as the length of the first phrase ($length in bars)

increase_weight_of_start = 1 #there is a slight preference for a phrase start so increase the weights of $\boldsymbol{S_1}$ rules

}

}

#Phrase label loop:

If $state = ***$S_1$***{

look back to previous states for which $phrase_label has not been updated.

If the following labels match then update $phrase_label

If $state = ***Then*** $phrase_label =

Rule Number	**$state (i.e. phrase part) combination**	**$phrase_label**	**Explanation**
PL.1	S_0 S_1 S_2^+ S_3^+ S_4^+ S_5 S_6	Complete phrase	#All phrase parts follow in sequence
PL.2	(S_3 ||S_5) S_6(through rest) S_1	PE and PS separated	#A rest between PE and PS
PL.3	(S_3^+ ||S_5) S_6 S_1	Confirmed phrase boundary	#PS follows PE
PL.4	(S_3^+ S_5 S_6) ¬ S_1	Contradictory, weakened PE	#PE not followed by clear PS, ($S_3^+$$S_5$ S_6) not followed by S_1
PL.5	S_3^+ S_4^+ S_5 S_6	Fulfilled expectation	#PE expected and achieved
PL.6	S_2 ¬ (S_3^+|| S_4^+|| S_5|| S_6)	Unfulfilled expectation: state 2 not followed by states 3-6	#PE expected but not achieved. Contradictory features and therefore phrase parts
PL.7	(S_1 follows (S_3 or S_5)) not preceded by S_2	Unprepared phrase boundary	#Unprepared phrase boundary; relatively weak
PL. 8	S_1 S_1 S_1^+ (each through repeats)	Sequence	#Several exact or inexact repeats
Comparing different parses			
PL.8	S_1 S_1 (S_3^+ || S_5)	Overlapping phrases	#PS precedes 'previous' PE
PL.9	S_1 on same $note as S_3^+ || S_5	Elided phrase boundary	#PS coincides with PE

If the combination in the list of states recorded does not exist in this table:

If this is not the only parse for the section **and** if the others do have entries in the table: **this parse has the least weight.**

elsif this is not the only parse for the section **and** if the others also do not have entries in the table **or** if this is the only parse: **this area is ambiguous.**

}

}

III. Concluding and presenting the outcome at end of the input:

```
        Print "analysis of online results"                          #For a summary of the 'online' results
Foreach $note   {
        Print $Label, $Location, $Category, $State and $no_of_rules_applied
        }
        Print "analysis of retrospective results"                   #For a 'retrospective' analysis
Foreach $note {
        If $Category = O, $Category = C                             #Change all to $Category = C
        Print $Label, $Location, $Category, $State and $no_of_rules_applied
}
```

Chapter 15

Application of the new concepts and methods to test pieces: Finale

15.1 Introduction

The analysis of the case-study pieces, and of other approaches to 'the phrase', lead to the identification of musical features, phrase parts, phrase types and the relations between them (chapters 10-14). In this chapter, these ideas are 'tested' using test pieces of the same type used in the case-studies (from the same eras, genres, and instrumental and vocal forces). The tables of test pieces and recordings are given in appendix 15.1.

Here, the test pieces are analysed and characterised according to their features (presented in feature graphs in appendix 15.2, like those of chapter 10) as defined in chapters 12-13 and more formally by applying the rule base in the form of an algorithm (chapter 14). These analyses are discussed and compared with the results of analysis of performance contours and an 'expert' analyst's annotation of scores, and, where available, with published discussions of the pieces.

The rule base and algorithm (chapter 14) are not the sole 'outcomes' of the preceding work, but rather a formalised summary of some aspects, particularly, the relationship between the different features and the phrase parts identified by participants and analysts and identifiable in performance contours. There are currently no data sets that could be used automatically for its testing without substantial interpretation from the user and reliance on several other modules each with its own difficulties. Therefore, the algorithm is currently not implemented and is instead applied manually. For implementation, the method of identification of the musical features in some cases (such as pitch jumps) would be clear, while for others there would be more difficulties (such as for 'inexact repeats') as has been encountered in other studies (discussed in chapter 12).

Each application consists of two stages: I. Initialising and II. Running the input. In the first the texture of the piece, where the excerpt begins (start of piece or not), and time and bar position, are set. In the 'running the input' stage, the rules are applied note-by-note, ending with labelling of the phrase parts. The application of the algorithm is discussed in detail for the first piece. There is much repetition when applying the rules to the subsequent pieces so for these pieces highlights are given in appendix 15.2.

While acknowledging the difficulties in deriving phrase parts from tempo and intensity contours (such as the tempo contours of the Brahms, chapter 10), they were shown to be helpful, and are used again. For reasons of space, the results of these analyses are summarised in the discussion of each piece (section 15.2).

The expert analyst is Dr. Dean Sutcliffe, a Reader in Eighteenth-Century Music and Director of Studies in Music at St Catherine's College, Cambridge. He specialises in analytical approaches to Baroque and Classical music and undertook the annotation of the test pieces. Dr. Sutcliffe teaches an analysis course which concentrates on Rothstein's theories (chapter 9) as the method for phrase analysis.[143]

Dr. Sutcliffe was given the pieces for a week to annotate the 'phrasing' and the beginning of the expectation of the end of the phrase. The examples were prepared in the same way as reported in chapters 2 and 6. The piece order was: Bach Suite TP[144], Mozart Sonata TP, Brahms TP, Bach Cantata TP, Chopin TP and Mozart Aria TPI and Mozart Aria TP II. Sutcliffe annotated phrase arcs, beginning and ending with circled notes (PS and PE), and an asterisk (*) marked the beginning of the expectation (EOP) (appendix 15.3 shows the annotations). After Dr. Sutcliffe had annotated the pieces, an interview was conducted.

During the interview Dr. Sutcliffe was asked to define 'the phrase' and to describe how one should be identified. Following Rothstein, he explained that a phrase concludes with cadential activity and that harmonic action has to have taken place (usually to a V or replacement). Dr. Sutcliffe was also asked for a definition of what the beginning of the expectation of the end is and how it should be identified. He explained that this is the chord before the dominant (the dominant preparation). The following questions were also asked:

---

143 Many thanks to Dr. Dean Sutcliffe for his insightful annotations and conversations.

144 TP stands for test piece

Did you take any notice of text?	No
Did you play the pieces?	Some
What was governing your phrase decisions: each part individually, one part, or the whole texture?	The top part because these pieces were mainly homophonic
Were you certain about every marking?	Yes, except where he marked a "?" [rare]
Is there ever a possibility of more than one interpretation?	Yes, but not if one is following Rothstein's theory [which he was]
Would you expect to see these phrases 'reflected' in performance characteristics e.g. tempo and dynamics?	Yes, through 'marginal' (or slight) tempo changes, mild agogic touches, slight weakening of intensity. The effects on performance for these pieces would be small. In other styles the effect may be larger.
How long can a phrase be?	Quite a few minutes[145] if there is no explicit pre-cadential activity.

Different people, even analysts, may arrive at different phrasing interpretations of the same piece (chapter 10). As Dr. Sutcliffe agreed, rather than giving a definitive phrase structure, these annotations are therefore to be taken as one possibility.

The comparison of the phrases indicated by: the performance features, the analyst and the musical features and their combinations (as formalised in chapter 14) followed the methods described in chapter 10. Following the discussions of each of the test pieces, an overall comparison between Dr. Sutcliffe's annotations and the annotations following the manual implementation of the algorithm is given.

15.2 The Test Pieces

15.2.1. Mozart, K.283 II Andante

Mozart's earliest surviving piano sonatas K. 279-83, dating from about 1775-7, are not works of great individuality; their general style is akin to that of J.C. Bach's sonatas (Sadie, 1982, p. 45). Here, the Andante of K. 283 is analysed.

The phrasing seems simple in comparison to many of the other pieces studied in this work, with repetitions, changes in motive, cadential and voice-leading progressions and resolutions, and rests. Despite this simplicity, there seem to be several phrasing options.

[145] This is very long in comparison to the lengths discussed in most psychological approaches.

15.2.1.2 Applying the Algorithm

I. Initialising:
A. Solo and Accompaniment,
B. Start of piece ($current_state = ***$S_1$***)
C. Set Time = 0, Beat =1, bar = 1

II. Running the input
As the excerpt starts at the beginning of the piece, there is an assumption that this is also a PS, so $current_state = ***$S_1$***. This is the case for all of the following test pieces. From this state, the next possible states are: ***S_2*** (beginning of the end of the phrase, the rules starting with 2.), ***S_3*** (arrival at resolution, the rules starting with 3.) and ***S_5*** (end of end, the rules starting with 5.). Therefore, although there is a pitch jump (rule 6.3, State ***S_6***) between the two quavers of beat 3, this is ignored. In the second bar there is a (5-4-)3-2-1 descent. However, this is not supported harmonically so this rule does not succeed either. In chapter 12 it was suggested that the strongest voice-leading progressions to signal the end of the phrase are those that descend to $\hat{1}$. However, there are others, as in the case of bar 2, such as the ascent to $\hat{1}$ which could be included in the definition of voice-leading. In bar 2 there is a cadence. The position given here for the EOP is the 6-4 chord: the V is clear in the bass and has already had some preparation. The $current_state changes to ***S_2***. The cadence resolves on bar 3. The bass note is the $\hat{3}$ not the $\hat{1}$ - this weakness of the resolution means that rule 3.1.4 (moving to $state=***$S_3$***) succeeds ($current_state = ***S_3***). The resolution coincides with the varied repetition of the opening (rule 1.1.2, moving to $state= ***$S_1$***), returning the $current_state to ***$S_1$***.

The next EOP occurs on bar 4.25, signalled by the cadential progression, implicit voice-leading, and a change in texture (rule 2.1 moving to $state= ***$S_2$***, rule 2.2, moving to $state= ***$S_2$*** and rule 2.3, moving to $state= ***$S_2$***), ($current_state = ***S_2***). This time the PE resolution is stronger (on bar 4.75) with $\hat{1}$ in the bass and melody than the previous PE (rule 3.1.1). This is followed by a rest (rule 6.1, moving to ***S_6***, $current_state = ***$S_6$***) and a PS signalled by change in motive (rule 1.2, moving to ***S_1***, $current_state = ***$S_1$***). In addition, the first phrase established a two bar phrase as the preferred length (rule I.C.3. bar number). This combination of features is stronger than those of bars 2-3. This difference in strength between the two PEs is not represented in Sutcliffe's markings, (the notation did not require this, section 15.2.1.3), it is seen clearly in the performance contours – for example, the tempo contours show a much larger *ritardando* than in bars 2-3 (section 15.2.1.3).

The counters from the initialisation stage therefore now favour phrases of 2 or 4 bars (I.C.3 bar number) starting on the first beat of the bar (I.C.2 beat number).

The next EOP is on 6.5 (rules 2.1, moving to $state= ***$S_2$*** and rules 2.2, moving to $state= ***$S_2$***, $current_state = ***$S_2$***) and the phrase cadences on G, signalling the PE (rule 3.1, moving to \$state= S_3, \$current_state = S_3). The PE is again followed by a rest in the right hand (rule 6.1, to S_6, \$current_state = S_6). The PE however, is weakened by the resolution of the right hand on to $\hat{3}$ rather than the $\hat{1}$ and the immediate continuation of the left hand.

The next PS (bar 7) is, as in bar 6, signalled by a change in motive (rule 1.2, \$current_state = S_1) and is followed by the EOP (bar 8, rules 2.1, moving to \$state= S_2 and rules 2.2, moving to \$state= S_2, \$current_state = S_2) and the PE (bar 8.75, rule 3.1, moving to \$state= S_3, \$current_state = S_3 and rule 6.1, moving to S_6,\$current_state = S_6).

For Sutcliffe bars 9 to the end form one phrase (section 15.2.1.3). However, the algorithm would label some additional shorter phrases. Bars 9 and 10 are a theme and its repeat (rule 1.1.5, \$current_state = S_1). There is small harmonic motion from V to V/V and both bars end with a rest (rule 3.1.4 \$current_state = S_3, rule 6.2, \$current_state = S_6). Bars 11.25 – 12.25 and 12.5 - 13 are also a theme and its repeat having harmonic motion from V-I (rule 1.1.5, \$current_state = S_2 , rule 3.1.4 \$current_state = S_3).

The phrase_label loop would label the PE and PS in bars 4-5 as the strongest and bars 2-3 and 9-10 as elided phrases. Table 15.2.1.5 summarizes the positions that would be identified by the algorithm.

15.2.1.3 Analyst's Annotations and performance contours

Sutcliff's annotations of this piece emphasise various possible relationships between PEs and PSs. In bars 4-5, 6-7 and 8-9, there is a rest in the melody between the PE and PS. In bars 4-5 there is a rest in the melody, and accompaniment, whilst in bars 6 and 8 there are linking semiquavers in the accompaniment. The first phrase ends on 2.625 and the next starts on 3. The relationship here is similar to the one in bar 4 of the Bach Prelude TP and is further discussed in section 15.2.2.3. The un-annotated notes between the phrases can be seen as beginning as an after-beat but turn into an anacrusis for the next phrase. Sutcliffe marks three EOPs before the PE in bar 14. As in other pieces, the cadential preparation begins and then the music 'backs up' before the final resolution forming a very long phrase (table 15.2.1.5). Within this phrase, there is a fermata (bar 13.25 and 27.25 in the repeat) that was omitted in the score provided to the analyst. Sutcliffe's phrase continues straight through it, and for in the interview he explained that the fermata does not affect phrase structure.

The tempo contours of the various performances are very similar; the beat lengthenings are primarily at the ends of bars 4, 13 and 14 (graph 15.2.1.3.1). There are also smaller, and sometimes more gradual, lengthenings (such as in bars 6, and 8). The whole opening section analysed in this study is repeated in all the performances and the repeats have beat lengthenings in the same areas though in some performances there are slight differences in location or degree of lengthening.

Graph 15.2.1.3.1 showing the tempo contours for the performances of Mozart Sonata K. 283 bars 1-14 and the repeat

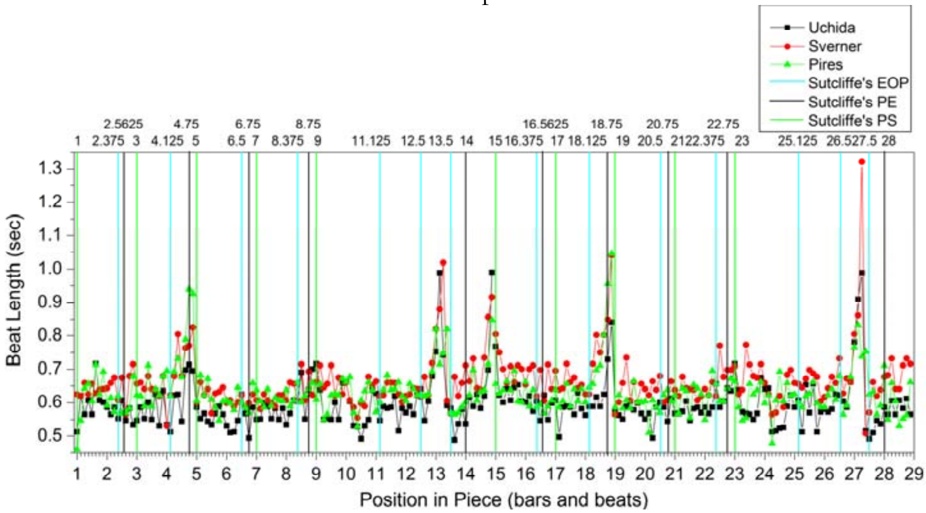


Table 15.2.1.3.1 summarises the intensity reduction of the performance contours and shows that, on the whole, they are also very similar.

Table 15.2.1.3.1: Bars including intensity reduction (Bold–greater)			
Bar number	**Uchida**	**Sverner**	**Pires**
4	**/**	**/**	**/**
8	/	/	/
10			
11			
13	**/**	**/**	**/**
14		**/**	**/**
Repeat:			
4 (18)	**/**	**/**	**/**
8 (22)	/	/	/
10 (24)		/	
11 (25)	/	/	/
12 (26)			/
13 (27)	**/**	**/**	**/**
14 (28)		**/**	**/**

15.2.1.4 Features and the various phrase interpretations

The feature graph (graph 15.2.1, appendix 15.2) shows five positions with at least four features and these coincide with Sutcliffe's PS. All involve either following a cadential progression, voice-leading repeats or changes in motives and some also coincide with the four-bar template and the position in the bar in which the first phrase started. The combinations here were found to coincide with PSs in the case-studies and therefore do so here. In several cases, positions with a smaller number of features precede these including mainly the first arrival on the resolution of a cadential or voice-leading progression. These would indicate PEs in the algorithm and coincide with Sutcliffe's PEs. The beginning of the voice-leading and cadential progressions in turn precedes these, signalling the EOP. The gap between PE and PS may, as in the Bach Prelude TP, be considered a link (section 15.2.2).

The annotations resulting from the application of the algorithm are generally similar to Sutcliffe's. However, especially in the second half, there are positions that would be annotated by the algorithm but were not chosen by Sutcliffe. These are weaker than the main phrases (and would be annotated as such), but the results of the listening study (chapters 3 and 10) indicate that some listeners might identify such positions.

As table 15.2.1.5 shows, some of Sutcliffe's PEs coincide with the greatest beat lengthenings and others do not. Of the latter, sometimes the nearest beat lengthening is relatively far (such as bar 2 and its equivalent 16) but others are closer or are part of a *ritardando* (such as bars 8 and 18) and fall between the PE and the following PS. When there is a *ritardando* towards the PE (as marked by Sutcliffe), it begins very soon after Sutcliffe's EOP. Many of the lower intensity positions in the performances coincide with the gap between Sutcliffe's PE and PS but rarely with the PE itself. Here, there is often a thinner texture or a rest. The diminuendi begin close to Sutcliffe's marked EOP.

Table 15.2.1.5: Comparisons of Algorithm's and Sutcliffe's annotations and interpretations of the performance contours							
Algorithm			Sutcliffe			Performances	
PS	EOP	PE	PS	EOP	PE	Beat-lengthening	Intensity Reduction
1 (S)	2.5	3 (W)	1	2.375	2.5625		
3 (W)	4.25	4.75 (S)	3	4.25	4.75	4 (S)	4 (S)
5 (S)	6.5	6.75 (W)	5	6.5	6.75	6 (1st time)	
7 (W)	8.375	8.75 (W)	7	8.375	8.75	8	8
9 (W)	9.5	9.75/10 (VW)	9	11.25		9 (2nd time)	10
10 (VW)	10.5	10.75/11 (VW)					11
11 (W)	11.5	11.625 (VW)					
11.75 (VW)	12.25	12.375 (VW)		12.5			12
12.5 (VW)	13	13.125 (VW)		13.5		13 (S)	13 (S)
13.375 (VW)	13.75	14.5 (S)			14	14 (S)	14 (S)
S = strong, W = weak, VW = very weak							

The same analyses were carried out on the rest of the test pieces. For reasons of space, the applications of the algorithm are given in appendix 15.2 and summaries of the performance analyses are provided here.

15.2.2 Bach, Suite for unaccompanied 'cello No. 2 Prelude, BWV 1008

15.2.2.1 The piece

One motive and its variant form the basis of the opening of this Prelude (the score is given in appendix 15.1). The features (and algorithm) indicate that several of the new PSs are signalled by the transposed repetitions of the motive (graph 15.2.2, appendix 15.2). The relatively clear phrase positions (indicated in all of the performance contours), are also preceded by voice-leading descents (explicit or implicit). Therefore, in some cases, such as with the motivic repetitions, the PSs are identifiable primarily retrospectively, in other cases, such as when there is clearly preparing voice-leading, the PSs are identifiable predictable, and sometimes, when both are present, they are both predicted and confirmed retrospectively.

15.2.2.2 Analyst's Annotations and performance contours

Sutcliffe annotated three phrases (appendix 15.1), each with an EOP, the second and third are elided (table 15.2.2.3). Areas 4.5-4.833 (and 24.166-25) are not included under arcs; they are transitions, a prolonged anacrusis. If these notes had been connected to a phrase it would have been the subsequent one. However, the parallelism with the opening is overriding, indicating the importance of both parallelism of bar position and harmonic motion. For Sutcliffe the EOP of the second phrase is relatively weak; chord VII is weaker than V. Sutcliffe's last phrase is long; there is a large scale descending sequence but the patterning is less clear. He explained that he had decided on the location of the PE and PS in retrospect.

Sutcliffe explains that he usually identified the position of the EOP by the position of the initial dominant preparation. This results in different lengths between EOP and PE, ranging here between 0.333 and 2.0 bars.

Overall Sutcliffe's annotations coincide more closely with the intensity contours than with tempo contours of the performances (table 15.2.2.3).

15.2.2.3 Features and the various phrase interpretations

Sutcliffe's annotations generally coincide with the PSs identified using the features (graph 15.2.2, appendix 15.2). As in the Mozart Sonata TP (section 15.2.1.3 above), Sutcliffe's separation between the PE and the PS in bars 4-5 and 24 highlights a case not covered by the phrase 'parts' developed in this study but resembles Rothstein's 'lead-in' (chapter 9). The phrase's resolution has been reached (I on bar 4), indicating a clear PE, and there is a repetition (bar 5), indicating a clear PS. A prolongation-based extension of phrase ends is included in the rules, however, the quavers joining the D in bar 4 and the B*b* in bar 5 are a link, or an upbeat to the next phrase. If this were polyphonic, the three quavers could be part of the accompaniment with the 'solo' returning at the start of bar 5, as in bars 6-7 of the Mozart Sonata TP (appendix 15.1).

The long (or main) phrases in this piece that would have been identified by the algorithm are the same as those identified by Sutcliffe and are very long (up to 13 bars). However, the listener is helped by more short-term punctuations marked by descents and inexact repeats of the opening. Listeners' responses to the case-study pieces suggest that some would identify these shorter units as phrases. This algorithm reflects both types of interpretation, placing more weight on the longer phrases (i.e. positions selected both for long and short phrases).

Table 15.2.2.3: Comparisons of Algorithm's and Sutcliffe's annotations and interpretations of the performance contours							
Algorithm			**Sutcliffe**			**Performances**	
PS	EOP	PE	PS	EOP	PE	Beat-lengthening	Intensity Reduction*
1 (S)	1.75	2 (VW)	1	3.666			
2 (VW)	2.75	3 (VW)					
3 (VW)	3.75	4.5-4.875 (S)			4.333		4.333-4.666
5 (S)	5.75	6 (VW)	5				
6 (VW)	6.76	7 (W)					
7 (W)	7.75	8 (VW)					
8 (VW)	8.75	9 (W)					
9 (W)	12	13 (S)		12. 666	13	12.666, 12.866, 13	13
13 (S)	14.75	15 (W)	13			13	
15 (W)	17.75	18 (W)					
18 (W)	18.75	19 (W)					
19 (W)	19.75	20 (W)					
20 (W)	23.75	24 (S)		22	24	24	23.666-24
* some of the common positions							

15.2.3 Bach, Cantata BWV 140 Wachet Auf, ruft uns die stimme, No. 4 *Zion hört die Wächter singen*

15.2.3.1 The piece

In this movement, for tenor (a *cantus firmus*), strings and continuo, the phrases of the 'solo' and 'accompanying' parts seems clear, but they do not always coincide (chapters 2 and 10). As is usual for music with a *cantus firmus*, the other parts are much 'busier'. This piece is one of 'an interesting class of examples in the history of music, where melodies are used in a way that they can hardly be directly perceived. During the Baroque era particularly, many works … were written as artful elaborations of a hymn. The hymn itself, called the *cantus firmus*, appeared in one of the parts at a very slow pace, surrounded by much faster moving and elaborated parts … A listener ignorant of this construction will naturally have his/her attention directed towards the faster parts, which allow the perception of various melodic-rhythmic patterns, whereas the listener won't at all detect the underlying *cantus firmus*, because it exceeds the limit of the psychological present to a high degree. The listener will instead experience the single tones of the *cantus firmus* as providing a harmonic foundation for the upper parts (which it in fact *also* is)' (Gabrielsson 1993, p. 97). The clearest features of the *cantus firmus*' phrasing are the very long rests (with each new entry starting a new text line). All but one of the entries of the *cantus firmus* start and end either on $\hat{1}$ or $\hat{5}$ and do not travel far. Many of the entries also have explicit voice-leading descents at their end, except bars 56-7 and 63-4.

In the orchestral exposition the possible phrasings are relatively clear; each phrase ends with preparatory features, cadential progressions of different strengths, or voice-leading that resolve especially upwards to the $\hat{1}$ or $\hat{5}$, and short boundary features (rests) separate them. There are very few phrase boundary areas that coincide exactly. In some cases, such as bars 18-19, one part (in this case the vocal one) seems to 'complete' the phrase of the other, but in most cases they overlap.

15.2.3.2 Analyst's Annotations and performance contours

Sutcliffe only marked the violin voice (the top line in the score) and the marked arcs most directly coincide with it (appendix 15.1). He explained that these relate to the whole texture but are dominated by the orchestral parts. For the most part, the annotations coincide with the rests in the instrumental parts. However, the whole passage of bars 6.875-17.5, and those like it, is annotated as one phrase although it contains two rests, providing evidence that rests alone do not determine phrasing.

Most of the PEs identified by Sutcliffe coincide with peaks in the tempo contours (table 15.2.3.3). However, many of the beat lengthenings in the tempo contours do not coincide with Sutcliffe's markings, possibly because he only annotates the top line. It seems that the 'accompanying' violin part determines the phrasing for Sutcliffe but is not the only determining part in the performances.

All of the PEs identified by Sutcliffe coincide with some minima in the intensity contour. However, there are large differences in degree, the extent to which the minima are preceded by diminuendi, and their lengths. Overall, there seems to be coincidence between the ends of the text lines and intensity minima. Although Sutcliffe, and Gabrielsson (1993), treat the vocal line as secondary in performances of such pieces, the intensity contours closely reflect the phrases of the vocal lines. In chapters 7 and 10 it was found that the interpretation of the performance contours was most problematic in pieces with many, polyphonic voices. Therefore, the comparison tables for such pieces, do not include the results of the performance contours.

15.2.3.3 Features and the various phrase interpretations

Unlike most of the other test pieces, in this piece only some of the positions with the largest number of features coincide with Sutcliffe's PS markings; bars 2, 4, 12 and 17 (table 15.2.3.3 and graph 15.2.3, appendix 15.2). Sutcliffe does not mark 7.875, 8.875 or 23. On the other hand, for most of the performers, most of the time, these areas do coincide with beat lengthenings and intensity drops. Sutcliffe does mark 6.875, which has only two new features, but the cadential progression features from the previous beat also apply here. Like in many of the pieces, there is often separation between the phrase end and phrase start. In most cases the separation is of a quaver or more rest. However, a rest last a quaver or more does not necessarily mean a PS.

The previous studies (chapter 10) indicate that there would be at least some listeners who would annotate each vocal interjection as a phrase while others would annotate some together (especially within bars 56-70). Some would annotate only according to the violin part and others would combine the two. Sutcliffe's responses are related only to the violin part. However, the performance contours indicate that the vocal line is also important. The algorithm includes the possibility of such a relationship between parts. It would suggest two groups of possible overall phrase structures: one vocally and one instrumentally based. The case-study responses showed that even though listeners may intend to respond to a particular instrumental or vocal part, they usually end up responding to more than one part, and the algorithm reflects this.

Table 15.2.3.3: Comparisons of Algorithm's and Sutcliffe's annotations [1]						
Algorithm				**Sutcliffe**		
PS		**PE**				
Orchestral	**Vocal**	**Orchestral**	**Vocal**	**PS**	**EOP**	**PE**
1		2.625		1	1.75	2.625
2.875		4.625		2.875	3.75	4.625
4.875		6.625		4.875	6.75	6.625
6.875		7.625		6.875	11, 11.75[2]	
7.875		8.625				
8.875		12.5				12.5
12.875	13.5	14.625	15.5	12.875	14	14.625
14.875		16.625		14.875	16	16.625
16.875	16.75	17.625	19	16.875		
17.875		18.625				
19.375	20.25	21.125			20.75	21.125
			22			
21.875		23.625		21.875	22.75	23.625
23.875		24.625		23.875	24.75	25.625
25.875		27.625		25.875	27.25	27.625
27.875		28.625		27.875		
28.875		29.625			32, 32.75[2]	
29.875		33.5				33.5
33.875	34.5	35.625		33.875		35.625
			36.5			
35.875		37.625		35.875		37.625
37.875	37.75	38.625		37.875		
38.875		39.625	40			
40.375	41.25	42.375			Same as	42.125

			43		equivalent earlier positions	
42.875		44.625		42.875		44.625
44.875		45.625		44.875		50.5
45.875		46.625				
46.875		50.5				
50.875	50.75	50.625		50.875		
51.875		51.625	52			
	53.25		54.5			(53.5-53.625)
54.375		56.125		54.375	55.625, 57.75[3]	
56.375	56	58.125	57			58.625
58.375		59.125		58.375		
59.375		60.125				
60.375		63			62	63
63.875		65.625		63.875	Same as equivalent earlier positions	
	63.25		64			
	65.75		66.5			65.625
66.375		68.125		66.375		68.125
68.375		69.125		68.375		
	60.25		70			
69.375		70.125				
70.375		74			71.75, 72.5, 73.5[2]	74

[1] As in the other pieces with the same texture only PSs and PEs are presented.
[2] Two or more EOP per phrase. For example, in bar 11, the cadence begins on 11, is thwarted and begins again at the end of the bar, the resolution arrives in the next bar.

15.2.4 Mozart, Don Giovanni, *Deh, Vieni alla finestra,* Canzonetta

15.2.4.1 The piece

This aria has a three-part texture for voice, mandolin and strings. As with the Bach Cantata TP, and unlike many solo arias, the mandolin has a clear melody of its own, sometimes with a phrase structure that is a little different from that of the voice. Also as with the Bach Cantata TP, the 'accompanying' mandolin part opens the aria, with an exposition of its melody. During the rest of the movement, the mandolin has repetitions and developments of these opening bars – sometimes the theme is more 'solo-like' and sometimes it is more in the style of a 'traditional' accompaniment.

15.2.4.2 Analyst's Annotations and performance contours

Sutcliffe mainly annotates the vocal part (table 15.2.4.2) and annotates the mandolin part when the vocal one has long rests. He explained that his annotations applied to all parts. Though he did not use the text, his phrases coincide with the text lines and all are separated by rests. Within the first text line, there is also a crotchet rest (coinciding with the comma, bar 6). This is longer than the rest separating this line from the next (bar 8). His fourth phrase starts on 12.25 because of the voice-leading in the accompaniment A-F#-E. He describes the end as peculiar - not a proper phrase; there is a post-cadential prolongation but without a "proper" approach.

Several PEs identified by Sutcliffe coincide with peaks in the beat-length. There are other positions in which peaks in the tempo contours fall between the EOP and the PE and yet others in which peaks that do not coincide with either of these positions. However, these often fall soon after the PE marks, often lengthening the 'gap' between the phrases. Several of the PE positions identified by Sutcliffe coincide with intensity minima. However, the minima more often follow Sutcliffe's mark. This is because there are often PEs in the mandolin part after those in the vocal parts. Many of Sutcliffe's EOP marks coincide with the peaks in intensity.

15.2.4.3 Features and the various phrase interpretations

There is a relationship between the positions with the greatest number of features (graph 15.2.4, appendix 15.2) and the positions Sutcliffe annotates as PSs. However, this is not completely clear from the presentation in the graphs as there, for example, the resolution of the cadence is marked at the location at which it first occurs (e.g. bar 14) rather than over the area (lasting until 14.5 in the violins). The features are spread over several notes while actually relating to the same PE, PS pair. Most of the EOP marks annotated by Sutcliffe coincide with the beginning of a cadential or voice-leading progression falls between this and the PE.

Two sets of phrases are identified with the algorithm, most of which overlap. The vocal phrases coincide with those chosen by Sutcliffe. Some listeners however, may refer more to the mandolin part.

Table 15.2.4.2: Comparisons of Algorithm's and Sutcliffe's annotations						
Algorithm - first part				**Sutcliffe**		
PS		**PE**		**PS**	**EOP**	**PE**
Orchestral	**Vocal**	**Orchestral**	**Vocal**			
1	4.833	4.5	6.333	1	2	4.5
5	6.833	8.916	8.5	4.833	6	8.5
9	8.833	14.5	12	5.833	10	12
				12.333	13	14.5
15 (secondary to voice)	14.833	18	18	14.833	17	18
18	18.33	24.5	22	18.833	21	22
25	24.833	28.916	28.5	24.833	26	28.5
29	29	32	32	29	31.333	32
32		34.5		32	33	34.5
35	34.833	38	38.833	34.833	37	38
39	39	44	42	39	41	42

15.2.5 Mozart, Don Giovanni, *Madamina*

15.2.5.1 The piece

In this aria for voice and orchestra, though there is a 'melody' in the accompaniment, it is less immediately clear than the melodies in the other such pieces discussed in this study. The vocal part varies from interjections to longer, more melodic segments.

The algorithm would annotate short, weak phrases in the vocal part signalled primarily by the rests, stronger ones when there is clear harmonic expectation and resolution and yet stronger ones when the vocal line coincides with the accompaniment. The responses to the vocal pieces in the listening studies (chapters 3 and 10) indicate that some listeners may also annotate the individual vocal interjections and phrases, others would annotate according to the accompaniment and still others would give a combination of the two. Sutcliffe annotates long phrases that coincide with those that this algorithm would assign to the accompaniment.

15.2.5.2 Analyst's Annotations and performance contours

Unlike the other pieces, Sutcliffe annotates phrases across the accompaniment and vocal parts. Like for the other pieces he explains that his annotation is for all the parts together. Unlike in the Mozart Aria TP I (section 12.2.4), Sutcliffe's annotations do not coincide with the text lines or even the vocal part specifically. Instead, they follow the accompaniment's harmonic progressions – from tonic to tonic. Sutcliffe does not mark an EOP for the last phrase. He explains that this section is post-cadential and may not be a phrase at all (table 15.2.5.2).

Table 15.2.5.2: Comparisons of Algorithm's and Sutcliffe's annotations								
Algorithm						**Sutcliffe**		
Vocal			**Orchestral**					
PS	**EOP**	**PE**	**PS**	**EOP**	**PE**	**PS**	**EOP**	**PE**
2.75 (S)	-	3.25	1	10		1	10	
4.25	4.5	5.25						
5.75	6.75	7.25						
7.75	-	10.25						
10.75	12.5	13			13			13
13.5	15.5	16 (S)	13.25	14.25	16	13.5	14.25	16
			16.25	19	20	16.25	19	20
18.5 (S)	-	20.25	20.25	23	25			
22.5	-	24.25						
26.5	-	28.25	24.25	27	28	24.25	27	28
28.75	32.5	33	28	31.25	33	28	31.25	33
34.5	(34.5)	35	33.25	36		33.25		
36.5	(36.5)	37 (S)			37		-	37 post-cadential?

In the performance contours, extreme lengthenings coincide with fermatas in bars 28 and 29. Relative to this, all the other tempo variations are very limited and there are few places for which all performers coincide. There are several areas in which all parts have intensity minima, some of which coincide with the analysts' annotations (such as around bar 24, or the two pauses of bars 28 and 29).

15.2.5.3 Features and the various phrase interpretations

For the most part there seems less coincidence between the features (graph 15.2.5, appendix 15.2) and the PSs marked by Sutcliffe than in some other pieces. This is mainly because some of the features, and therefore the algorithm, take into consideration the vocal part, while Sutcliffe's annotations do not. The algorithm, however, would also annotate several of the phrase parts similar to those annotated by Sutcliffe if the same voice is considered.

15.2.6 Chopin, Nocturne Op. 9 No. 1 in B flat minor

15.2.6.1 The piece

Op. 9 No. 1 is one of Chopin's first published nocturnes.[146] Chopin's op. 9 effectively formalized the tradition. In most of his nocturnes, an ornamental aria alternates with a sequentially developing, tension-building theme. Nocturne Op. 9 No. 1 almost immediately introduces a style of decoration seldom encountered in those of Chopin's contemporaries, and never so near the beginning of a work (Samson, 1994, pp. 46). The irregularity of the rhythmic patterns in, for example bars 1 – 4, is an aspect of Chopin's ornamentation style that continues to find increasingly varied expression in later works, including Op. 27 No. 2 (Samson, 1994, pp. 46-7).

For Samson, in this Nocturne at 'the level of the phrase, there is the immediate ornamental variation of the opening two-bar shape. There are three such variations, increasingly elaborate, in the Nocturne as a whole, and they are all expressive outgrowths of the melody rather than superimposed decorations. This , too, is characteristic. Finally, at the level of the paragraph, the phrase structure is typically decepetive. Only at bars 15-18, as the second theme approaches, is the innocent ear made aware that the barring of the opening theme implied an iambic (weak-strong) rather than trochaic (strong-weak) rhythmic pattern. Again, this charactistics play on symmetries and asymmetries within an apparently simple phrase structure is as characteristic of Chopin as it is of Mozart' (Samson, 1996, p. 100-101).

The vocal, Italianate origin of the nocturne encouraged a phrase organisation in duple lengths – four, eight, sixteen, thirty-two bars (Rothstein, 1988, p. 118). Chopin adopted the formal and rhythmic conventions of his time, and with them the problems of phrase rhythm. He seems to have conceived of the latter as, above all, a problem of continuity, and he set about solving it by finding means (many of them traditional) to enhance melodic continuity which became increasingly elaborate (Rothstein, 1988, p. 141).

15.2.6.2 Analyst's Annotations and Performance Contours

Again, Sutcliffe annotates only the top part explaining that this phrasing applies to the accompaniment as well. Sutcliffe's annotations (table 15.2.6.3 and appendix 15.1) highlight the great variety of phrase lengths and relationships between EOP

---

[146] The character piece for piano (such as the Brahms chapter 10) was already well established by the time Chopin began composing them in the late 1820s. By the 1820s there was some consistency in the nocturne, especially among composers associated with John Field. The idea of vocal imitation whether of French *romance* or Italian aria was central, facilitated by the development of the sustaining pedal, enabling now architypal arpeggiations supporting the ornamental melody.

and PE. The phrasing coincides with the stronger phrases identified by the algorithm.

In the performances, beat lengthenings occur both on individual notes and on several that follow each other (table 15.2.6.3). All the PE areas around the four-bar phrase lengths are marked by phrase-final lengthening. Many of the PEs identified by Sutcliffe coincide with lengthenings and there are several cases of lengthening between the EOPs and the PEs identified by Sutcliffe. There are also several lengthenings that do not coincide with either (though most of these do not occur for all of the performances).

Many of the PEs identified by Sutcliffe coincide with lower intensity in the different performances. There are other areas with lower intensity (often a smaller difference) that do not coincide with any of Sutcliffe's markings.

15.2.6.3 Features and the various phrase interpretations

Most positions with several features coincide with PSs identified by Sutcliffe (2.5, 4.5, 8.5, 10.5, 12.5, see features graph 15.2.6, appendix 15.2). However, bar 17 has as many features as bar 2 but no phrase part is identified by Sutcliffe in the latter. In all the performances this bar is part of a *ritardando* and *diminuendo*. One of the features leading up to bar 17 is a cadential progression, which is not resolved here. Instead, there is an inexact repeat. This section is therefore an extension rather than a PE.

In several cases the left and right hand resolutions do not exactly coincide. For example, in bar 2 the harmonic progression resolves onto the tonic B*b* on the first beat. The melody line sounds a d*b* and this only resolves onto the b*b* two beats later (with the confirming arpeggio in the accompaniment). Therefore, the beginning of the end can be seen as occurring on beat one, while the end of the end occurs on beat three.

The algorithm "annotations" of the longer phrases would coincide with Sutcliffe's. For this piece, any lengthening coincides with at the PS or PE indicated by musical features.

Table 15.2.6.3: Comparisons of Algorithm's and Sutcliffe's annotations and interpretations of the performance contours							
Algorithm bold - relatively strong			**Sutcliffe**			**Performances** Bold - relatively strong	
PS	EOP	PE	PS	EOP	PE	Beat-lengthening	Intensity Reduction
0.5	**1.75**	**2.333**	0.5	1.75	2.333		2.333
2.5	**3.666**	**4.333**	2.5	3.666	4.333	3 – 5	4.333
4.5	**7.5**	**8-8.333**	4.5	7.5	8-8.333	**7.83-8.333**	5.5,6-6.5,8.333
8.5	**9.666**	**10.333**	8.5	9.666	10.333		10.333
10.5	**11.666**	**12.333**	10.5	11.666	12.333	10 – 11	12.333
12.5	13.5	14.333	12.5			12-13.5	
14.5	15.5	16				15.5 (Hewitt)	
16	16.5	17					
17	17.5	**18-19**		17.5	18-18.333	**17.5 – 19**	17.5

15.2.7 Brahms, Intermezzo Op. 116 no. 2

15.2.7.1 The piece

This piece is in ternary ABA' form and this study looks at the A section. The phrases identified by the algorithm would be in two groups: four long ones relying on the preparation for, and resolution of, cadences as well as repeats and long notes, and the shorter ones relying on long notes and repetitions (table 15.2.7.3).

15.2.7.2 Analyst's Annotations and Performance Contours

Again there are different relationships between the PEs and PSs in Sutcliffe's annotations (table 15.2.7.3). One pair occurs from one note to the next (bb. 13-14), another is separated by a bar (bb. 4-5) and in another, the PE occurs on the last note of the right hand of one bar and the PS occurs on the first beat two bars later (bb. 8-10).

As in the Mozart Sonata TP and Bach Prelude TP (sections 15.2.1 and 15.2.2) some notes are not included under phrase arcs in Sutcliffe's annotation. Unlike the Bach Prelude TP, where the quavers were moving away from the arrival, leading to the next PS, here, these notes form an extension of the previous phrase. In bars 9 and 18 they are an exact repeat an octave lower of the closing bar of the previous phrase. Following the algorithm, these would have been included as part of PE prolongation and confirmation.

Although bar 13 is a variant of bar 4, Sutcliffe annotates the two differently. He explains that the melodic line it different; it is more active the second time. The first time there is more weight on the first beat of the bar. This indicates that although harmonic and voice-leading motion may dominate, other factors, such as rhythm, play a part in phrase structure identification.

The tempo and the intensity contours are similar especially in comparison to the contours for the Brahms Intermezzo (chapter 7) though there are differences of degree. Most of the beat-lengthenings coincide with Sutcliffe's annotations (table 15.2.7.3).

15.2.7.3 Features and the various phrase interpretations

The positions with the most and 'strong' features are 5, 10 and 14 (see graph 15.2.7, appendix 15.2). These coincide with Sutcliffe's PSs. In the intensity contours, several of the minima fall between the PE and the PS but only some are separated by lengthening in the tempo contours. The long phrases identified by the algorithm are similar to those identified by Sutcliffe. There would be a small discrepancy however, between Sutcliffe's markings and those of the algorithm. Sutcliffe distinguishes between different PE positions in the variants of bar 4 while the algorithm would mark these all in a similar way.

Table 15.2.7.3: Comparisons

Algorithm			Sutcliffe			Performances	
PS	EOP	PE	PS	EOP	PE	Beat-lengthening	Intensity Reduction
1	3.333	4.333	1	3.333	4	1, 2	4.333
5	7.333	9.666	5	7.333	8.333-666	7, 9	10
10	12.5	13.333	10	12.5	13.333		12-14
14	15	18	14	15	17.333	15-16, 18	16-19

15.3 General comparison between annotations by analyst and algorithm

The above analysis has discussed each piece and compared the performances, analyst's and algorithm annotations in detail. In order to provide a comparison between the analyst's annotations (A) and those of the algorithm (R), the evaluation scheme used in natural language, PARSEVAL is applied. This is used by Bod (2002), allowing a comparison with the results of the purely memory based approach (see chapter 8.5). This method combines two measures: precision and recall. Precision calculates the number of phrases identified by the algorithm and also by the analyst (i.e. "correct" phrases in R), divided by the total number of phrases identified by R. Recall calculates the total number of phrases identified by the algorithm and also by the analyst (i.e. "correct" phrases in R) divided by the total number of phrases identified by A. Beyond testing for the "percent correct", these measures together test both for too many phrase assignments (precision) and too few (recall) and a good result is a high score in both.

$$\text{Precision} = \frac{\text{“correct” phrases in } R}{\text{Phrases in } R}$$

$$\text{Recall} = \frac{\text{“correct” phrases in } R}{\text{Phrases in } A}$$

These two are combined into a single measure of performance the F-score (Manning and Schütze, 1999).

$$\text{F-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In this test, one “correct” parse has to be assumed, and the test is whether or not the algorithm identifies it. Sutcliffe’s annotations are taken as the “correct parse” though, there can be other interpretations (some of which are indicated by the performance contours discussed above). Furthermore, as discussed in chapters 11-14, at some level, different notes can be identified as the same phrase part. In this test therefore, phrase areas (of two beats) are taken as “equivalent”. As the expert analyst did not provide different strengths with his annotations, only the strongest phrase parts identified by the algorithm are included in this calculation. In addition, when the music is polyphonic and the algorithm provides more than one output, the one relating to the part Sutcliffe annotated is taken. Following Bod (2002), the algorithm’s phrase identifications are taken as correct if PS and the following PE match those of the analyst. The results are calculated from the comparisons tables above and are shown in table 15.3.1. For a general comparison only, the results are shown alongside the results for the folksongs tested and reported in Bod (2002).

Table 15.3.1: Summary table showing results of PARSEVAL application

	Memory-base, Essen folk songs (Bod, 2002)	Algorithm,Combined Approach Test pieces, current study
Precision	0.77	0.85
Recall	0.86	0.97
F-score	0.81	0.91

The results show that, overall, the annotations of the analyst and the algorithm are similar. There is one piece (the Bach Cantata TP) that for which the interpretation for the same part does not match Sutcliffe’s completely. In this case, the algorithm would identify shorter phrases than identified by Sutcliffe, at least on one level, explaining the lower precision than recall score (as discussed in section 15.2.3.3).

15.4 Summary of the analyses of the test pieces

The analyses of the test pieces used the feature analysis and the rule base and algorithm developed on the basis of the case-study pieces for the purpose of phrase-part identification. The phrasing results were compared with performance contours, and an analyst's phrase markings. This comprehensive analysis of phrasing provides an assessment of the factors that affect phrasing and their relative strength on the one hand, and the sources of variability in phrase interpretation on the other. In general, using the algorithm it was possible to identify phrase parts and phrase combinations and these were largely confirmed by the analyst and often reflected in performance contours.

15.4.1 Performance contours, analyst's annotations and the algorithm

In each piece there are several positions for which either an increase in beat-length or a decrease in intensity or both in at least one performance contour coincide with positions identified by the analyst and the algorithm. In general, the analyst and algorithm identify positions that are emphasised in at least one of the performance contours. There are also changes in performance features in the performance contours, possibly indicating non-phrasing aspects, that do not coincide with the phrase parts marked by the analyst or the algorithm (chapter 10).

15.4.2 The analyst, features and the algorithm

In general the analyst's phrase marks coincide with those identified using the features. The positions with the highest number of features, usually also have a phrase mark annotated by the analyst and, conversely, the different positions of the EOP, PE and PS features coincide with the analyst's annotations. Areas with a larger number of features include strong features (cadential progressions, voice-leading, changes and repetitions) that do not occur elsewhere, while for in the rest of the positions usually have 'weaker' features.

In each piece, there are several positions for which the analyst's markings coincide with the phrase parts identified by the algorithm. However there are occasions in which phrase parts identified by the algorithm are not identified by the analyst. This often stems from the analyst's approach of not annotating phrasing of more than a single voice. In addition, the analyst usually annotates only "confirmed" PEs and PSs (i.e. positions at which the phrase has reached its end and a PS follows immediately or almost immediately), while the algorithm takes into account listeners' responses patterns, which included also weaker phrase parts.

15.4.3 Multiple voices

Sutcliffe's annotation of the multi-voice or polyphonic pieces indicates that the 'solo' or vocal line does not necessarily dictate the phrasing. For Sutcliffe, the phrases are primarily dictated by the harmonic structure that is most clearly expressed in the accompaniment and there is no need for them to coincide exactly with the vocal line.

The listening results indicated that this is sometimes the case for listeners too (such as in the Bach Passion, chapter 10). However, there are some listeners that even in such pieces do respond either to both the vocal and orchestral parts, or only the vocal line. For example, most listeners and musicians annotate only the vocal line in response to such pieces as the Schubert songs. The performance contours show that the vocal line often seems to influence overall positions of changes in tempo, even in the orchestral part. As indicated by listeners' responses to the case-study pieces (chapter 10), the vocal line does play an important rôle and is therefore not omitted from consideration by the algorithm. The algorithm relies both on many features that are most explicit in the orchestral parts and others that are in the vocal parts, therefore reflecting the phrase parts that could be identified in one, the other or both.

15.4.4 Phrase Parts

The phrase parts identified using the algorithm match most of those identified by the analyst and those highlighted in the performance contours. The phrase part combinations seem to reflect the different possibilities suggested by the analyst well, including phrases that end on one note and start on the next, elided phrases, separated phrase ends and starts, or expectations whose resolutions are delayed. Together, the rules of the algorithm seem to be helpful in describing not only the complete strong phrases but also the weaker or partial phrases that analysts and performers respond to.

The discussions of the Mozart Sonata TP, the Bach Suite TP and Brahms TP (sections 15.2.1-2 and 7) indicated that there is an additional phrase part. The algorithm already includes a 'gap' state (chapters 12-14) usually indicated by a rest (rules 6.1-2 and 6.4-5). If such a gap occurs, in one voice, another voice may still be playing in another part, sometimes 'linking' the two otherwise separate phrases (such as in the Mozart Aria in the case-study pieces, see chapter 10, section 10.6.1.2). In some cases, this 'link' filling the 'gap' between phrases occurs all in the same part(s) as in these test pieces. These 'links' are usually an 'anacrusis' for the following PS and so would be identified by such features as leading back to a repeat. It therefore seems that a further 'gap' type should be included in this phrase part: a linking, anacrustic phrase-part. For some analysts, as demonstrated by Sutcliffe, these sections may be excluded from the phrase structure. The aims of this algorithm include the identification of the 'function' of each phrase part, therefore, these sections are included among the phrase labels.

15.4.5 The rule base and algorithm

The comparison of the algorithm with the analyst's annotations and performance contours and, in previous chapters, with listeners' and musicians' responses, suggests that the algorithm, based on feature-identification, gives a representation of the different phrase-part interpretations. The algorithm is versatile and can be "set" to give some of the possibilities, depending on characteristic of the type of "listener" being "modelled". For example, to model an analyst like Sutcliffe, the algorithm would concentrate on the rules that consider the harmonic structure and texture as a whole at the expense of the details of the solo vocal part. Conversely, modelling some of the musicians who annotated the score 'for performance' (chapter 6) the algorithm could give details of all the voices and the stronger and weaker phrases.

Chapter 16

Summary:
Coda

Various perspectives on the musical phrase

The notion of musical phrase has been previously investigated and used within the disciplines of musical analysis and psychology. It was further developed in this study through experimental and analytical approaches that identified its essential components and their variability in relation to different musical contexts.

The phrase consists of phrase parts and forms a unit that endows the music with organisation and order for the analyst, performer and listener (chapter 1). It does so through numerous means such as elements that create: expectation, frustration and resolution (usually discussed in music analysis, chapters 1 and 9); similarity, contrast, and variation (used in gestalt-based psychological approaches, chapters 1 and 8); motion and arrival (mentioned in some approaches to performance, chapter 7); and boundary, breath and rest (mentioned in all approaches).

The distribution of these different elements between the separate disciplines (and therefore also methodologies) means that, in many cases, different elements are treated separately, thus automatically removing the possibility of investigating the relationships between them. This study used a combined approach and shows that by bringing together these different elements and by discussing them in view of various theories pertaining to the respective disciplines and approaches, a more comprehensive and complete investigation of the phrase is achieved.

This provision of organisation and order has several aspects relating to the internal structure of the phrase and the relation between phrases. Within the phrase, the location of phrase parts is signalled by the occurrence of certain musical features and their combination (chapters 10-13). Different features bring with them different cognitive loads, with the phrase ends and starts having lower cognitive loads than the middle, as revealed by the click experiments (chapter 5). This is one of the ways in which the musical phrase is comparable to segments in other media such as language.

Feature combinations determine phrase types, which relate to each other differently and influence the relationship between phrases. For example, a front-heavy phrase can be followed by an end-heavy one, forming an antecedent-consequent phrase relation between the two.

Phrase identification

The phrases investigated and referred to in the analyses in this study were identified by a varied population from different perspectives: listeners (with different degrees of musical experience who carried out phrase identification in MIDI and two different sets of performances, chapters 3-5 and 10); score readers (phrase identification in scores as if preparing for performance, chapters 2 and 6); music analysts and theorists, (carrying out general music analyses and a specific expert analyst who annotated the test pieces, chapters 9, 10 and 15), music psychologists (discussing phrasing, segmentation and grouping, chapter 8), and performances (through the analysis of the performance features of tempo and intensity changes, chapter 7).

The study concentrated on three groups of excerpts (introductory-study, case-study and test-pieces, chapters 2, 3-4 and 15 respectively) selected for their range of eras, instrumental combinations, and textures, and their musical and particularly phrase characteristics.

In terms of the population investigated and the media of presentation, this seems to be the broadest study on phrase identification so far. The results indicate that musical experience did not systematically affect responses. The differences in response that were identified seem to indicate that musical training enables only 'clearer' responses to 'subtleties' or 'irregularities' of phrasing (chapter 1) and not substantial differences between groups; indeed, there were differences in responses among the trained musicians.

The results from all the above perspectives suggest that there are a number of 'candidate' phrase positions which form two categories: those chosen by the majority of participants in most cases and those identified by some participants less frequently, even though they have similar musical training. For the latter category, even the individual listeners are not always self-consistent in their identification of the phrase positions. This contrast in frequency of identification indicates a difference in relative importance among phrase positions (chapters 3 and 4).

The responses to the scores, MIDI, and performances of the same pieces show that a number of positions were identified in all media. For other positions, the responses varied both within and between media. Most interestingly, all the positions identified in score annotation and in responses to performances were also identified in response to the MIDI (chapters 3 and 4). The differences between, for example, responses to performances and to MIDI were in terms of proportion, not position. The structure, as suggested by these proportions, seems to have remained with the listeners and affected their responses in the subsequent listenings to the second recording (chapters 3 and 10). Performance characteristics therefore do not seem to alter the 'candidate' positions but do seem to affect their strengths and this stays with the listener. Similarly, previous familiarity with the piece seems to affect phrase structure identification in some cases (chapter 3).

These results indicate that musical features alone (i.e. not including 'performance features') are very important in phrase identification even by untrained listeners. In the context of responding to MIDI, performances and the score, all participants seemed to have intuitions about what a phrase is and what they were 'looking for' (chapter 3).

These results also suggest that the long term 'learning' effects do seem to be important in phrase identification. However, short-term 'learning', hearing the same piece several times consecutively, does not seem to have a systematic influence on responses (this observation was also made by Deliège, chapter 8).

Phrases and their features

Having recorded and analysed the phrase identifications using a number of different methodologies (from music psychology, natural language processing, music performance, music analysis and new ones developed here – the combined approach), the primary method of analysis of the features in relation to phrases identified in this study was as follows. The features were defined and identified on the basis of music theoretical and psychological criteria and were located through analysis. The frequency of individual and combined features, and of their co-occurrence with phrases identified in this study was calculated and documented numerically and graphically (chapters 10-11).

This paved the way for:

1) The identification of the relative 'importance' of the different features and feature combinations in the phrase. For example, a perfect cadence resolving on the tonic note in the bass, and a descent through $\hat{3}$-$\hat{2}$-$\hat{1}$ in the melody followed by a repetition, always coincides with a phrase end, indicating a relatively 'important' features combination. On the other hand, other features, such as relatively large pitch intervals or long notes occur more frequently and do not always co-occur with phrase parts. In general, features occurring less frequently are more important for phrase identification and vice versa.

2) The observation that the relative importance can change depending on context, either within a piece or in terms of genre. For example, in vocal music rests become more important than in purely instrumental music.

3) The classification of features according to three categories:
Scope of presence:

a. Instantaneous; one note to the next, or longer,
b. Area, with a clear beginning, or
c. Area, with emphasis towards its end.

Scope of impact:

a. Instantaneous,
b. Retrospective, or
c. Combined predictive and retrospective, and

Function or result:

a. Instantaneous,
b. New beginning,
c. Expecting/achieving the end.

This classification reflects the characteristics of the features studied here and indicates that it may be possible to draw equivalences between some features. As seen in this study, in some genres, features that are in general not important become so in specific ones. The features categorisation described here may therefore be useful if further features are to be investigated, particularly in the context of different genres or other music with different specific musical features which have similar scopes of presence and impact and functions or results.

4) The identification of different phrase parts, phrase types and the relation between phrases. This allows for a better understanding of: a) the ways in which the phrase is internally 'organised' and how the phrases 'organise' the structure of the piece, b) the internal structure of the phrase and the possible effect this has on phrase identification, particularly the rôle of expectation. This study identified the following phrase parts:

- Phrase Start,
- Beginning of the expectation of the end,
- Beginning of the end,
- Arrival at the resolution
- End of the end (end of the resolution)

each one having its associated (more or less) indicative features. Phrases can be front-heavy (characterised by features such as repetitions) or end-heavy (goal directed phrases, leading to cadences). c) The understanding that where 'ambiguity' arises this results from the combinations of features that may suggest more than one phrase structure. The phrase part combinations lead to the identification of different phrase types and phrase type combinations. The 'musical surface', features, phrase parts, phrase types, and phrase type combinations are hierarchically structured.

The characterisation of phrases and phrase parts using features concluded from the analysis of the case-study pieces constitutes the basis for a formal rule base.

This can be implemented when a larger corpus of annotated music becomes available (chapter 14). The rule base was then applied to test pieces of the same genre as the case-study pieces. It is not intended to provide 'one solution' rather to reflect different possible moment-by-moment identifications of phrase parts and phrase combinations.

Phrases and beyond

This study indicates that:
1. Musical and performance features in the context of phrasing help to organise incoming information online.
2. Phrases can contain starts, expectation, frustration and realisation, and are sometimes separated by boundaries.
3. The cognitive load is affected by phrase structure and plays a rôle in its identification.
4. Memory and the related connection between musical features are crucial to the identification of phrases, phrase parts and relationships between phrases.
5. The relationship between 'the whole' and 'the part' both in terms of phrases and in terms of different voices can change between phrases.
6. The metric structure provides a frame in relation to which the harmonic features and phrases in general are organised.
7. Phrase identification is not a result of training.

Many of these conclusions may be comparable with (but not necessarily analogous to) aspects of the perception of language (and other media).

The results of this study show that it is neither surprising that phrasing is mentioned so often in so many different approaches within music theory and analysis, music psychology and computational musicology, nor that its definition is often left vague. It is important in all of these approaches and disciplines, and is multi-facetted in its contributing features and constituents, its nature, and its implications for other and more general aspects of music and music perception.

Moreover, this study of musical phrasing, by embracing many different aspects that are often explored within the fields of music theory, analysis and psychology, has shown them to be relevant to musical phrasing, such as in contributing to its identification, including expectation, or to other aspects, such as emotional, or proto-emotional, responses (chapter 1). This, in combination with the conclusion that phrase identification is not an ability limited to highly trained musicians, indicates not only that musical phrasing is common to those who participate in this kind of music (as listeners, performers or analysts) but also that musical phrasing is key to music perception and analysis in general.

Samenvatting

Wat draagt bij tot de perceptie van muzikale frasen in westerse klassieke muziek?

De muzikale frase

Bestaat ze? Voor wie? Waar?Wat? Wanneer? Waarom?

Deze vaak gebruikte term, zo intuïtief begrepen door vele muzikanten, roept verschillende associaties op met de terminologie van vele disciplines zoals muziek, psychologie, fysiologie en linguïstiek. Het wezen van de frase blijft echter duister.

Bestaat ze?

Eén van de voornaamste doelstellingen van deze studie was om vast te stellen in welke mate er gemeenschappelijke ideeën bestaan over het wezen van de frase, haar beschrijving, identificatie en functie.

Voor wie?

Een andere doelstelling was de identificatie van de types populaties voor wie deze entiteit relevant is. Deze identificatie geschiedde door het onderzoeken van (i) verbale en muzikale responsen gegeven door luisteraars met verschillende muzikale ervaring; (ii) partituren geannoteerd door muzikanten alsof ter voorbereiding van een uitvoering; (iii) karakteristieken van de uitvoeringen van publiekelijk beschikbare opnames; (iv) analyses door muziekpsychologen, muziekanalisten and muziektheoretici; (v) muzikale analyse van stukken aan de hand van bepaalde aspecten (hieronder aangeduid als de 'gecombineerde analyse'). Deze populaties werden reeds voorheen bestudeerd, maar niet met dergelijke directe en gedetailleerde methodes. The resultaten geven aan dat gemeenschappelijke aspecten van de frase niet aangeleerd zijn; luisteraars met verschillende niveaus van muzikale training of zonder enige training beantwoordden op een gelijkaardige wijze de taken en vragen met betrekking tot frasen.

Waar?

Deze studie bespreekt voorbeelden van muzikale frasen van verschillende muzikale genres en media. Sommige van deze voorbeelden werden reeds onderzocht binnen de context van verschillende disciplines. De voorbeelden variëren van folk tot westerse klassieke muziek. De kern van deze studie is echter de toepassing van de gecombineerde analyse op acht stukken gebruikt als casestudy's, gevolgd door analyses van zeven teststukken, allemaal uit het westerse klassieke repertoire.

Eén van de vragen betreft de mate waarin de frasen en hun structuren duidelijk zijn 'van de partituur' (dwz. van de muzikale elementen die kunnen geïdentificeerd worden in de partituur), en in welke mate deze elementen enkel duidelijk zijn in een uitvoering. Door het gebruik van responsen op MIDI-uitvoeringen (van luisteraars met verschillende muzikale ervaring) en annotaties in de partituur (door muzikanten), werd vastgesteld dat gemeenschappelijke frasestructuren duidelijk zijn door hun muzikale elementen. Muzikale elementen worden gecombineerd en hebben bepaalde karakteristieken afhankelijk van hun context. Voorbeelden zijn: verschillende cadensen, relatief grote intervallen in toonhoogte of lange noten of rusten, herhalingen en veranderingen in textuur, motief, en harmonisch ritme.

De resultaten tonen verder aan dat bepaalde kenmerken van de uitvoering (veranderingen in tempo en intensiteit) ook een rol spelen. Tempo en intensiteit veranderen in opnamen en werden ook vergeleken met: 1) frasestructuur geïdentificeerd door analisten, personen die partituren annoteren en luisteraars naar MIDI-uitvoeringen en 2) de responsen van luisteraars op dezelfde opnamen. Deze geven aan dat dezelfde plaatsen in de muziek die naar voren kwamen in de uitvoering ook geïdentificeerd worden in de andere media, en dat de responsen van luisteraars op uitvoeringen in verband staan met de elementen van de uitvoering. De belangrijkste plaatsen geïdentificeerd in responsen op uitvoeringen en MIDI-uitvoeringen zijn dezelfde, maar er is een proportioneel verschil in het aantal van de responsen. Bovendien, wanneer één uitvoering gehoord werd, wordt haar frasestructuur blijkbaar herinnerd en beïnvloedt de identificatie van frasen in volgende uitvoeringen (laat 'voetafdrukken' na). Frasering zit blijkbaar fundamenteel 'in de muziek' en wordt geaccentueerd, verduidelijkt of verdoezeld door de uitvoering.

Frasering is voornamelijk bestudeerd binnen de context van monofone muziek. In deze studie wordt muziek van verschillende textuur bestudeerd. De resultaten van de gecombineerde analyse geven aan dat in polyfone muziek (b.v. melodie en begeleiding) er verschillen kunnen bestaan tussen de frasestructuur van verschillende delen; frasen in verschillende stemmen kunnen in elkaar overvloeien of elkaar vervolledigen, overlappen of samenvallen, en deze verschillen worden vaak geïdentificeerd door participanten. Dit geeft aan dat we zowel de individuele delen als de structuren van de frase van verschillende muzikale stemmen identificeren, en deze aanpassen in een meer algemene identificatie van conflicterende, complementaire en gelijkaardige frasestructuren.

Wat?

De term frase heeft verschillende synoniemen in de literatuur, die ook gebruikt worden door de deelnemers aan deze studie ("segment, unit, chunk, sentence", etc.). Muziekpsychologische en computerwetenschappelijke analyses concentreren op de identificatie van frasegrenzen, terwijl andere soorten analyses interne karakteristieken analyseren.

De hierboven besproken gecombineerde analyse geeft aan dat frasen bestaan uit enkele van de volgende delen: begin, begin van het einde (implicatie/verwachting), einde (initiële aankomst), verlenging (voortzetting van het einde), en einde van het einde (einde van de oplossing), en dat elk deel aangegeven wordt door specifieke muzikale elementen. Hoewel al deze verschillende delen aanwezig kunnen zijn, is hun aanwezigheid niet noodzakelijk opdat de frase geïdentificeerd, herkend of geïmpliceerd kan worden. De vermelde analyses onderzochten ook het relatieve belang van de verschillende delen van de frase, wat het karakter van de frase bepaalt, zoals beginzware en eindzware frasen, en mogelijke verhoudingen tussen de frasen, zoals antecedent-consequente fraseparen.

Voor al deze zaken lijkt de sleutel te liggen in de aanwezigheid of afwezigheid van muzikale elementen. Deze behoren tot verschillende categorieën, die een verschillende mate van aanwezigheid, impact en functie hebben. Sommige kunnen onmiddellijk zijn (vinden plaats, worden geïdentificeerd, en hebben een repercussie van de ene noot op de andere, zoals een grote interval in toonhoogte), sommige kunnen voorspellend zijn (vinden plaats over een gebied en scheppen verwachtingen, bijvoorbeeld ontwikkelende harmonische progressies zoals cadensen) en sommige kunnen retrospectief zijn (vinden opnieuw plaats over een gebied, maar geven hun belang achteraf bloot, zoals herhalingen). Verschillende elementen en combinaties van elementen vallen blijkbaar systematisch samen met verschillende aantallen responsen, geïdentificeerd door de gecombineerde analyse. Sommige kenmerken en hun combinatie vormen een sterke aanwijzing voor de aanwezigheid van een frase, terwijl dat bij andere minder het geval is. De eerste groep komt meer frequent voor dan de tweede. Afhankelijk van de muzikale context (zoals genre, instrumentale combinaties of lokale context) verkrijgen gemeenschappelijke elementen een groter belang. Bovendien benadrukt deze gecombineerde analyse de interdependentie van de muzikale elementen; verschillende combinaties van harmonische, metrische en toonhoogte-structuren, kunnen bijvoorbeeld sterkere of zwakkere eindes van frasen vormen. De combinaties van elementen en frasedelen kunnen van die aard zijn dat meer dan één mogelijkheid kan ontstaan (wat soms resulteert in 'ambiguïteit').

Hoewel de verhoudingen tussen de elementen, frasedelen en frasen complex zijn en afhankelijk van verschillende parameters, zijn ze geformaliseerd in een algoritme. In tegenstelling tot andere algoritmes, is de bedoeling hier om het proces weer te geven van de identificatie van frasen, inclusief de 'zwakke' frasen, door deelnemers, en het aangeven van alternatieve mogelijkheden, gebruik makend van het concept van muzikale elementen ontwikkeld aan de hand van de gecombineerde analyse. Dit algoritme resulteert in duidelijk en consistente frasestructuren, en kan in de toekomst geïmplementeerd worden voor de studie van een meer uitgebreid assortiment muziek.

Wanneer?

Deze elementen en combinaties van elementen resulteren blijkbaar in kandidaat-plaatsen voor het begin, einde en de interne delen van frasen. Sommige plaatsen worden gekozen door de meerderheid van de deelnemers terwijl andere minder frequent worden geïdentificeerd. Deze laatste plaatsen vallen samen met zwakkere kenmerken en dus delen van de frase. Ze zouden waarschijnlijk genegeerd worden in 'eenvoudige annotaties' zoals in de Essen Folk Song Collection, maar blijken een integraal en belangrijk deel te vormen van het proces van luisteren, uitvoeren en analyseren.

Bovendien, door de hierboven besproken gecombineerde analyse en door het gebruik van klik-studies, en in tegenstelling tot vroegere psychologische studies, wordt het duidelijk dat delen van een frase vaak geïdentificeerd worden over een bepaalde periode, en niet op specifieke noten.

Waarom?

Deze studie geeft aan dat de frase zowel een organiserend als een georganiseerde eenheid is (in dit opzicht gelijkaardig aan een zin in taal) die in verhouding staat tot geheugen, adem, en fysieke beweging. Het geeft structuur, kader, orde en referentie en werkt in op structuren van verschillende types (b.v. ritmische structuren). Haar lengte wordt beschreven als constant. Echter, de resultaten van de gecombineerde analyse geven aan dat er een grote variabiliteit bestaat in lengte van frase. De identificatie van deze eenheden kan bijdragen tot herinnering en tot de vergelijking tussen gelijkaardige frasen en tot de meer algemene structurering en herinnering van de muziek. De frase helpt in het volgen van beweging of progressie vanaf een begin, het aankomen bij een bestemming of het terugkeren. Muzikale implicaties en daarom verwachtingen, spelen blijkbaar een belangrijke rol in deze progressie. Bovendien, uit de manier waarop de frase, haar muzikale kenmerken en karakteristieken gebruikt worden en hun frequent voorkomen in besprekingen van muzikale analyses, percepties en uitvoeringen, blijkt dat ze essentieel is voor ons vermogen het soort muziek hier bestudeerd te kunnen volgen.

References

Abecasis, D., Brochard, R., Granot, R.Y. and Drake, C. (2004), 'ERP exploration of metrical accents in isochronous auditory sequences', *in* Lipscomb, S.D., Ashley, R., Gjerdingen, R.O. and Webster, P. (eds.), *International Conference on Music Perception and Cognition*, Evanston IL.

Abrams, K. and Bever, T.G. (1969), 'Syntactic structure modifies attention during speech perception and recognition', *Quarterly Journal of Experimental Psychology*, 21, 280-290.

Ambrose, Z.P. (2004), 'Translation of Text of St. Matthew Passion'.

Baddeley, A.D. (1990), *Human memory: Theory and practice*, Hove, UK, Lawrence Erlbaum Associates Ltd.

Baron-Cohen, S. (2005a), 'Sex differences in mind', *Twentieth Annual Darwin College Lecture Series 2005: Conflict*, University of Cambridge.

Baron-Cohen, S. (2005b), 'The essential difference: The male and female brain', *Phi Kappa Phi Forum*, 85, 23-26.

Barra, D. (1983), *The dynamic performance: A performer's guide to musical expression and interpretation*, Englewood Cliffs, N.J., Prentice hall.

Bengtsson, I. and Gabrielsson, A. (1983), 'Analysis and synthesis of musical rhythm', *in* Sundberg, J. (ed.), *Studies of Music Performance*, Stockholm, Royal Swedish Academy of Music.

Bent, I. and Drabkin, W. (1987), *Analysis*, London, Macmillan Press.

Bent, I.D. and Pople, A. 'Analysis', *in* Macy, L. (ed.), *Grove Music Online*.

Berent, I. and Perfetti, C.A. (1993), 'An on-line method in studying music parsing', *Cognition*, 46, 203-222.

Blombach, A. (1987), 'Phrase and cadence: A study of terminology and definition', *Journal of Music Theory Pedagogy*, Volume 1, No. 2, 225-251.

Bod, R. (1993), 'Using an annotated language corpus as a virtual stochastic grammar', *in* Kaufmann, M. (ed.), *AAAI'93*, Menlo Park.

Bod, R. (1998), *Beyond grammar: An experience-based theory of language*, Stanford, CSLI Publications (distributed by Cambridge University Press).

Bod, R. (2001), 'Challenging the Gestalt principles in music', *Proceedings International Computer Music Conference (ICMC-2001)*, Havana, Cuba, extended version.

Bod, R. (2002), 'Memory-based models of melodic analysis: Challenging the Gestalt principles', *Journal of New Music Research*, 30, 1, 27-37.

Boersma, P. and Weenink, D. 'Praat', Institute of Phonetic Sciences, University of Amsterdam, Herengracht 338, 1016CG Amsterdam, The Netherlands.

Bolton, T.L. (1894), 'Rhythm', *American Journal of Psychology*, 6, 145-238.

Bond, Z.S. (1972), 'Phonological units in sentence perception', *Phonetica*, 25, 129-139.

Bregman, A. (1990), *Auditory scene analysis: the perceptual organization of sound.*, Cambridge, Mass, MIT Press.

Brown, M.J.E. and Sams, E. (1982), *The New Grove: Schubert*, Macmillan Publishers.

Burnett, J. (1972), *Brahms: a critical study*, London, Dent.

Cambouropoulos, E. (1998a), 'Musical parallelism and melodic segmentation', *Colloquium of Musical Informatics XII*, Gorizia, Italy.

Cambouropoulos, E. (1998b), 'Towards a general computational theory of musical structure ', *Faculty of Music and Department of Artificial Intelligence*, Edinburgh, University of Edinburgh.

Cambouropoulos, E. (2001), 'Melodic cue abstraction, similarity, and category formation: A formal model', *Music Perception*, 18, 3, 347-370.

Cambouropoulos, E. (2003), 'Musical pattern extraction for melodic structure', *Proceedings of the 5th Triennial ESCOM Conference 8th September 2003*, Hanover University Music and Drama, Germany.

Cambouropoulos, E. (2006), 'Musical parallelism and melodic segmentation: A computational approach. ' *Music Perception*, 23, 3, 249-267.

Caplin, W.E. (2004), 'The classical cadence: conceptions and misconceptions', *Journal of the American Musicological Society*, 57, 1, 51-118.

Carletta, J. (1996), 'Assessing agreement on classification tasks: the kappa statistic', *Computational Linguistics*, 22, 249-254.

Carletta, J., Isard, A., Isard, S., Kowtko, J.C., Doherty-Sneddon, G. and Anderson, A.H. (1997), 'The reliability of a dialogue structure coding scheme', *Computational Linguistics*, 23, 13-31.

Chaffin, R. and Imreh, G. (2002), 'Practicing perfection: Piano performance as expert memory', *Psychological Science*, 13, 4, 342-347.

Chailley, J. (1972), *"Tristan et Isolde" de Richard Wagner*, Paris, Leduc.

Charniak, E. (1996), 'Tree-bank grammars', *in* Church, K. and Gale, W. (eds.), *AAAI-96*, Menlo Park, Ca.

Chew, G. 'Articulation and Phrasing', *in* Macy, L. (ed.), *Grove Music Online*.

Clarke, E.F. (1988), 'Generative principles in music performance', *in* Sloboda, J.A. (ed.), *Generative Processes in Music: The psychology of performance, improvisation, and composition*, Oxford, England, Clarendon.

Clarke, E.F. (1989), 'The perception of expressive timing in music', *Psychological Research*, 51, 2-9.

Clarke, E.F. (1999), 'Rhythm and timing in music', *in* Deutsch, D. (ed.), *The Psychology of Music*, New York, Academic Press.

Clarke, E.F. and Baker-Short, C. (1986), 'The imitation of perceived rubato: A preliminary study', *Psychology of Music*, 15, 58 - 75.

Clifton, T. (1983), *Music as heard*, New Haven, Yale University Press.

Cohen, L. and Mehler, J. (1996), 'Click monitoring revisited: An online study of sentence comprehension', *Memory and Cognition*, 24, 1, 94-102.

Collins, M. (1999), 'Head-driven statistical models for natural language parsing', Pennsylvania, University of Pennsylvania, PA.

Cone, E.T. (1974), *The composer's voice*, Berkeley, Los Angeles and London, University of California Press.

Cone, E.T. (1998), "Am Meer' reconsidered: strophic, binary, or ternary?' *in* Newbould, B. (ed.), *Schubert studies*, Brookfield, Ashgate.

Cook, N. (1994), 'Perception: a perspective from music theory', *in* Aiello, R. and Sloboda, J. (eds.), *Musical Perceptions*, Oxford, OUP.

Cooper, G.W. and Meyer, L.B. (1960), *The rhythmic structure of tonal music*, Chicago, University of Chicago Press.

Cotard, C. (1895), *Tristan et Iseult, Essai d'analyse du drame et des Leitmotifs*, Paris, Fischbacher.

Couperin, F. (1772), 'Pièces de Clavecin. Livre III', Paris.

Court, R. (1986), *Essai sur les fondements anthropologiques des l'art*, Paris, Klinckseick.

Cutler, A. and Norris, D. (1979), 'Monitory sentence comprehension', *in* Cooper, W.E. and Walker, E.C.T. (eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, Hillsdale N.J., Erblaum.

Dahlhaus, C. (1990), *Studies on the origin of harmonic tonality*, Princeton University Press.

Damiani, A., Lorenzo, P.D., Maio, G.D. and Beardinelli, M.O. (2003), 'Recognising the composition style by themes selected form the "sonata" repertoire: a comparison between the mathematical modeling and the experimental psychological results', *in* Kopiez, R., Lehmann, A.C., Wolther, I. and Wolf, C. (eds.), *5th Triennial ESCOM Conference*, Hannover University of Music and Drama, Germany, Springer.

Davie, C.T. (1966), *Musical structure and design*, London, Dover Publications.

Davies, S. (1983), 'Attributing significance to unobvious musical relationships', *Journal of Music Theory*, 27.

Davis, S. (1966), 'Harmonic rhythm in Mozart's sonata form', *The Music Review*, XXVII, 25–43.

Deathridge, J. and Dahlhaus, C. (1984), *The New Grove: Wagner*, Macmillan Press.

Deliège, I. (1987), 'Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's Grouping Preference Rules', *Music Perception*, 4, 4, 325-360.

Deliège, I. (1989), 'A perceptual approach to contemporary musical forms', *in* S., M. and Deliège, I. (eds.), *Music and cognitive sciences*.

Deliège, I. (1998), 'Wagner "Alte Weise": Une approche perceptive', *Musicæ Scientiæ*, Special Issue, 63-90.

Deliège, I. and EL Ahmadi, A. (1990), 'Mechanisms of cue extraction in musical groupings: A study of perception on Sequenza VI for viola solo by L. Berio', *Psychology of Music*, 18, 1, 18-44.

Deutsch, D. (ed.) (1999), *The Psychology of Music*, Academic Press, London.

Di Eugenio, B. (2000), 'On the usage of Kappa to evaluate agreement on coding tasks', *In LREC2000: Proceedings of the Second International Conference on Language Resources and Evaluation*, Athens.

Di Eugenio, B. and Glass, M. (2004), 'The kappa statistic: a second look', *Computational Linguistics*, 30, 95-101.

Dorian, F. *The history of music in Performance: The art of musical interpretation from the Renaissance to our day.*

Dowling, W.J. (1984), 'Development of musical schemata in children's spontaneous singing', *in* Crozier, W.R. and Chapman, A.J. (eds.), *Cognitive Processes in the Perception of Art*, Amsterdam, North-Holland.

Dowling, W.J. and Harwood, D.L. (1986), *Music Cognition*, San Diego, Academic Press.

Drake, C. and Palmer, C. (1993), 'Accent structures in music performance', *Music Perception*, 10, 343-378.

Dunsby, J. (1981), *Structural ambiguity in Brahms: Analytical approaches to four works*, Michigan, UMI Research Press.

Dunsby, J. (1983), 'The multi-piece in Brahms: *Fantasien* Op. 116', *in* Pascall, R. (ed.), *Brahms: Biographical, Documentary and Analytical Studies*, Cambridge, CUP.

Dürr, W. (1982), 'Schubert's songs and their poetry: reflections on poetic aspects of song composition', *in* Eva Badura-Skoda and Branscombe, P. (eds.), Cambridge, CUP.

Edelman, G.M., Gall, W.E. and Cowan, W.M. (eds.) (1988), *Auditory function: neurobiological bases of hearing*, Wiley, New York.

Eysenck, M.W. and Keane, M.T. (1995), *Cognitive psychology: A student's handbook*, Hillsdale, Psychology Press.

Ferrand, M., Nelson, P. and Wiggins, G. (2002), 'A probabilistic model for melody segmentation', *Electronic Proceedings of the 2nd International Conference on Music and Artificial Intelligence (ICMAI 2002)*, Edinburgh, Scotland, UK.

Ferrand, M., Nelson, P. and Wiggins, G. (2003), 'Unsupervised learning of melodic segmentation: a memory-based approach', *Proceedings of the 5th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM5).*

Flores d'Arcais, G.B. (1978), 'The perception of complex sentences', *in* Levelt, W.J.M. and d'Arcais, G.B.F. (eds.), *Studies in perception of language*, Chichester, UK, Wiley.

Fodor and Bever (1965), 'The Psychological Reality of Linguistic Segments', *Journal of Verbal Learning and Verbal Behaviour*, 4, 414 – 420.

Fodor, J.A., Bever, T.G. and Garrett, M. (1974), *The psychology of language*, New York, McGraw-Hill.

Forte, A. (1998), 'A Schenkerian Reading of an Excerpt from Tristan und Isolde', *Musicæ Scientiæ*, Special Issue, 15-26.

Fraisse, P. (1978), 'Time and rhythm perception', *in* Friedman, E.C.C.M.P. (ed.), *Handbook of Perception*, New York, Academic Press.

Fraisse, P. (1982), 'Rhythm and Tempo', *in* Deutsch, D. (ed.), *The Psychology of Music*, New York, Academic Press.

Friberg, A. and Battel, G.U. (2002), 'Structural Communication (Chapter 13)', *in* Parncutt, R. and McPherson, G.E. (eds.), *The Science and psychology of music performance: Creative strategies for teaching and learning*, Oxford, OUP.

Friberg, A., Bresin, R., Frydén, L. and Sundberg, J. (1998), 'Musical punctuation on the microlevel: Automatic identification and performance of small melodic units', *Journal of New Music Research*, 27, 3, 271-292.

Friberg, A. and Sundberg, J. (1999), 'Does music performance allude to locomotion? A model of final *riterdandi* derived from measurements of stopping runners', *Journal of the Acoustical Society of America*, 105, 3, 1469-1484.

Frisch, W. (1981-2), 'Brahms, Developing Variation, and the Schoenberg Critical Tradition', *Nineteenth Century Music*, 5, 215-32.

Gabrielsson, A. (1987), 'Once again: The theme from Mozart's piano sonata in A major (K. 331). A comparison of five performances', *in* Gabrielsson, A. (ed.), *Action and perception in rhythm and music*, Stockholm, Royal Swedish Academy of Music.

Gabrielsson, A. (1988), 'Timing in music performance and its relations to music experience', *in* Sloboda, J.A. (ed.), *Generative Processes in Music: The psychology of Performance, Improvisation, and Composition*, Oxford, England, Clarendon.

Gabrielsson, A. (1993), 'The complexities of rhythm', *in* Tighe, T.J. and Dowling, W.J. (eds.), *Psychology and Music: The understanding of melody and rhythm*, Hillsdale, New Jersey, Lawrence Erlbaum Associates, Publishers.

Gabrielsson, A. (1999), 'The Performance of Music', *in* Deutsch, D. (ed.), *The Psychology of Music (second edition)*, California, Academic Press.

Garrett, M., Bever, T.G. and Fodor, J.A. (1966), 'The active use of grammar in speech perception', *Perception and Psychophysics*, 1, 30-32.

Grave, F., K. (1980), '"Rhythmic Harmony" in Mozart', *The Music Review*, XLI, 2, 87-102.

Greene, D.B. (1970), 'Schubert's "Winterreise": A Study in the Aesthetics of Mixed Media', *The Journal of Aesthetics and Art Criticism*, 29, 2, 181-193.

Gregory, A.H. (1978), 'Perception of clicks in music', *Perception and Psychophysics*, 24, 2, 171-174.

Handel, S. (1989), *Listening*, Cambridge, Mass, MIT Press.

Hartmann, A. (1932), 'Untersuchungen über das metrische Verhalten in musikalischen Interpretationsvarianten', *Archiv für die gesamte Psychologie*, 84, pp. 103-192.

Hauser, M.D., Chomsky, N. and Fitch, T. (2002), 'The faculty of language: what is it, who has it, and how did it evolve?' *Science*, 298, 1569-79.

Henderson, M.T. (1936), 'Rhythmic organisation in artistic piano performance', *in* Seashore, C.E. (ed.), *Objective analysis of musical performance, University of Iowa Studies in the Psychology of Music*, Iowa City, University of Iowa Press.

Hinson, M. (ed.), *Brahms: the shorter piano pieces*, Alfred.

Hirsch, I.J. (1959), 'Auditory Perception of Temporal Order', *The Journal of the Acoustical Society of America*, 31, 6, 759-767.

Hochberg, J. (1998a), 'Gestalt Theory', *in* Gregory, R.L. (ed.), *The oxford companion to the mind: the classic guide to the mind and its mysteries*, Oxford, OUP.

Hochberg, J. (1998b), 'The Oxford companion to the mind: The classic guide to the mind and its mysteries', *in* Gregory, R.L. (ed.), *The Oxford Companion to the mind*, Oxford, OUP.

Holmes, V.M. and Forster, K.I. (1970), 'Detection of extraneous signals during sentence recognition', *Perception and Psychophysics*, 7, 297-301.

Honing, H. (2001), 'From time to time: The representation of timing and tempo', *Computer Music Journal*, 35, 3, 50-61.

Honing, H. (2005), 'Is there a perception-based alternative to kinematic models of tempo rubato?' *Music Perception*, 23, 1, 79-85.

Honing, H. (2006), 'Evidence for tempo-specific timing in music using a web-based experimental setup', *Journal of Experimental Psychology: Human Perception and Performance*, 32, 3, 780-786.

Howell, P., West, R. and Cross, I. (eds.) (1991), *Representing Musical Structure*, Academic Press, London.

Huron, D. (1996), 'The melodic arch in Western folksongs', *Computing in Musicology*, 10, 3-23.

Imberty, M. (1998), 'Du vide à l'infini: homologies structurales repérées dans Tristan à partir du solo de cor anglais du III° acte', *Musicæ Scientiæ*, Special Issue, 91-116.

Jackendoff, R. (1992), 'Musical Processing and Musical effect ', *in* Jones, M.R. and Holleran, S. (eds.), *Cognitive Bases of musical communication*, Washington D.C., American Psychological Association.

James, W. (1890), *The principles of psychology*, New York, Holt.

Jones, M.R. (1992), 'Attending to musical events', *in* Jones, M.R. and Holleran, S. (eds.), *Cognitive bases of musical communication*, Washington D.C., American Psychological Association.

Jones, M.R. and Holleran, S. (1992), 'Cognitive bases of musical communication', *in* Jones, M.R. and Holleran, S. (eds.), *Cognitive bases of musical communication*, Washington D.C., American Psychological Association.

Jusczyk, P.W. and Krumhansl, C.L. (1993), 'Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure.' *Journal of Experimental Psychology: Human Perception and Performance*, 19, 3, 627-640.

Juslin, P.N., Friberg, A. and Bresin, R. (2001-2002), 'Toward a computational model of expression in performance: The GERM model', *Musicae Scientiae*, special issue, 63-122.

Juslin, P.N. and Sloboda, J. (eds.) (2001), *Music and Emotion: Theory and Research*, OUP, Oxford.

Kaminska, Z. and Mayer, P. (1993), 'Transformation, migration and restoration: Shades of illusion in the perception of music', *Contemporary Music Review*, 9, 1 & 2, 151-161.

Kellas, G., Ferraro, F.R. and Simpson, G.B. (1988), 'Lexical ambiguity and the time course of attention allocation in word recognition', *Journal of Experimental Psychology: Human Perception & Performance*, 14, 601-9.

Keller, H. (1956–7), 'A slip of Mozart's: Its analytical significance', *Tempo*, 42, 12–15.

Keller, H. (1957), 'Functional Analysis: Its pure application', *Music Review*, xviii, 202–6.

Keller, H. (1965), *Phrasing and Articulation: A contribution to a rhetoric of music with 152 musical examples*, London, Barrie and Rockliff.

Kerman, J. (1956), 'Opera as Symphonic Poem', *Opera as Drama*, Faber & Faber.

Kerman, J. (1980), 'How we got into analysis and how to get out', *Critical Inquiry*, 7, pp 311-31.

Kirkpatrick, R. (1984), *Interpreting Bach's Well-Tempered Clavier: A performer's discourse of method*, New Haven and London, Yale University Press.

Koch, H.C. (1787, 1983), *Introductory essay on composition*, New Haven and London, Yale University Press.

Koffka, K. (1935), *Principles of Gestalt psychology*, New York, Harcourt Brace.

Köhler, W. and Wallach, H. (1944), 'Figural after-effects: An investigation of visual processes', *Proceedings of the American Philosophical Society*, 88, 269-357.

Koniari, D., Predazzer, S. and Mélen, M. (2001), 'Categorization and Schematization Processes Used in Music Perception by 10- to 11-Year-Old Children', *Music Perception*, 18, 3, 297-324.

Kramer, J. (1993), 'Beyond Unity: Toward an Understanding of Musical Postmodernism', *in* Marvin, E.W. and Hermann, R. (eds.), *Concert Music, Rock and Jazz since 1945*, Rochester, NY, University of Rochester Press.

Kramer, J.D. (1988), *The time of music: New meanings, new temporalities, new listening strategies*, New York and London, Schirmer.

Krippendorff, K. (1980), *Content Analysis: An introduction to its methodology*, Beverly Hills, CA., Sage Publications.

Krumhansl, C. (1983), 'Perceptual Structures for Tonal Music', *Music Perception*, 1, 1, 28-62.

Krumhansl, C.L. (1990), *Cognitive foundations of musical pitch*, Department of Psychology, Cornel University, OUP.

Krumhansl, C.L. and Jusczyk, P.W. (1990), 'Infants' perception of phrase structure in music', *Psychological Science*, 1, 1, 70-73.

Krumhansl, C.L. and Kessler, E.J. (1982), 'Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys', *Psychological Review*, 89, 334–368.

Krumhansl, C.L. and Shepard, R.N. (1979), 'Quantification of the hierarchy of tonal functions within a diatonic context', *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579-594.

Kubovy, M. and Howard, F.P. (1976), 'Persistence of a pitch-segregating echoic memory', *Journal of Experimental Psychology: Human Perception & Performance*, 2, 531-537.

Labussière, A. (1992), 'Die alte Weise': une analyse sémiologique du solo de cor anglais du 3e Acte de Tristan et Isolde', *Analyse musicale*, 30-53.

Ladefoged, P. and Broadbent, D.E. (1960), 'Perception of sequence in auditory events', *Quarterly Journal of Experimental Psychology*, 12, 162-170.

Lampl, H. (1996), *Turning notes into music*, Lanham, Md. & London, The Scarecrow Press, Inc.

Landis, J.R. and Koch, G.G. (1977), 'The measurement of observer agreement for categorical data', *Biometrics*, 33, 159-174.

Lerdahl, F. (1998), 'Prolongational structure and schematic form in Tirstan's "Alte Weise"', *Musicæ Scientiæ*, Special Issue, 27-41.

Lerdahl, F. (2001), *Tonal Pitch Space*, Oxford, OUP.

Lerdahl, F. and Jackendoff, R. (1977), 'Toward a formal theory of tonal music', *Journal of Music Theory*, 21, 1, 111-171.

Lerdahl, F. and Jackendoff, R. (1987), *A Generative Theory of Tonal Music*, Cambridge, Massachusetts and London, England, MIT Press.

Lewin, D. (1986), 'Music Theory, Phenomenology, and Modes of Perception', *Music Perception*, 3, 4, 327-392.

London, J. (2004), *Hearing in Time*, Oxford, Oxford University Press.

Longuet-Higgins, H.C. and Lee, C.S. (1982), 'The perception of musical rhythms', *Perception*, 11, 2, 115-128.

Longuet-Higgins, H.C. and Lee, C.S. (1984), 'The rhythmic interpretation of monophonic music', *Music Perception*, 1, 4, 424-441.

Longuet-Higgins, H.C. and Lisle, E.R. (1989), 'Modelling musical cognition', *Contemporary Music Review*, 3, 15-27.

Longuet-Higgins, H.C. and Steedman, M.J. (1971, Republished 1987), 'On Interpreting Bach', *in* Meltzer, B. and Michie, D. (eds.), *Machine Intelligence*, Edinburgh, University of Edinburgh Press.

MacDonald, M. (1990), *Brahms*, London, Dent.

Macpherson, S. (1912), *Studies in phrasing and form*, London, Geo Barber.

Macy, L. 'Phrase', *in* Macy, L. (ed.), *Grove Music Online*.

Manning, C. and Schütze, H. (1999), *Foundations of Statistical Natural Language Processing*, Cambridge, MA, MIT Press.

Martinez, I.C. (2001a), 'La Prolongacion como un constituyente structural en la audicion musical atenta', *in* Shifres, F. (ed.), *Actas de la Primera Reunión Anual de SACCoM*, Buenos Aires.

Martinez, I.C. (2001b), 'The use of Prolongation in Music Attending', *Abstracts of the SMPC 2001 Conference*, Queens, Canada.

Martinez, I.C. (2002), 'Prolongation and Music Attending: A Click Localization Study', *International Conference on Music Perception and Cognition*, Proceedings of ICMPC7.

Matlin, M.W. (2003), *Cognition*, Wiley.

Mattheson (1737), 'On the Subdivision of Musical Language, 5th section', *Kern melodischer Wissenschaft (The essence of Melodic Science).*

McLaughlin, T. (1970), 'Music and Communication'.

Meredith, D., Wiggins, G.A. and Lemström, K. (2001), 'Pattern Induction and Matching in Polyphonic Music and Other Multidimensional Datasets ', *5th World Mulitconference on Systemics, Cybernetics and Informatics (SCI2001)*, Orlando, Florida.

Meyer, L., B. (1989), *Style and Music*, Philadelphia, University of Pennsylvania Press.

Meyer, L.B. (1956), *Emotion and Meaning in Music*, Chicago and London, The University of Chicago Press.

Meyer, L.B. (1973), *Explaining Music: Essays and Explorations*, Berkeley, University of California Press.

Meyer, L.B. (2001), 'Music and emotion: distinctions and uncertainties', *in* Juslin, P.N. and Sloboda, J.A. (eds.), *Music and Emotion: Theory and Research*, Oxford, OUP.

Miller, G.A. (1956), 'The magical number seven, plus or minus two: Some limits of our capacity for processing information', *Psychological Review*, 63, 81-97.

Moore, B.C.J. (1995), *Hearing*, San Diego, Academic Press.

Moore, G. (1975), *The Schubert Song Cycles with thoughts on performance*, London, Hamish Hamilton.

Morgan, R.P. (2003), 'The Concept of Unity and Musical Analysis', *Music Analysis*, 22, 1-2, 7 – 50.

Namba, S., Nakamura, T. and Kuwani, S. (1977), 'An analysis of piano performance (in Japanese)', *Studies in the Humanities and Social Sciences*, 25, 25-43.

Narmour, E. (1990), 'The Analysis and Cognition of Basic Melodic Structures: The Implication-Realisation Model'.

Narmour, E. (1992), 'The Analysis and Cognition of Melodic Complexity: The Implication-Realisation Model'.

Nattiez, J.-J. (1998), 'Le solo de cor anglais de Tristan und Isolde: essai d'analyse sémioloique tripartite', *Musicæ Scientiæ*, Special Issue, 43-62.

Negus, A. (1993), 'A musical commentary', *Tristan und Isolde, Wagner, Opera Guide 6*, London, Calder Publications.

Neumann, F. (1993), 'Performance Practices of the 17th and 18th Centuries '.

Newbould, B. (1977), 'A new analysis of Brahms's intermezzo in B minor, op.119 no.1', *Music Review*, xxxviii, 33-43.

Newman, E. (1961), *The Wagner Opera (Wagner Nights)*, Plimpton Press.

Ockelford, A. (Fall 2002), 'The magical number two, plus or minus one: some limits on our capacity for processing musical information', *Musicae Scientiae*, Vol VI, no. 2, 185-219.

Ottman, R. (1986), *Music for sight singing*, Englewood Cliffs, NJ, Prentice-Hall.

Palmer, C. (1997), 'Music Performance', *Annual Review of Psychology*, 48, 115-38.

Palmer, C. and Krumhansl, C.L. (1987a), 'Independent temporal and pitch structures in determination of musical phrases', *Journal of Experimental Psychology: Human Perception and Performance*, 13, 1, 116-126.

Palmer, C. and Krumhansl, C.L. (1987b), 'Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing and familiarity', *Perception and Psychophysics*, 41, 6, 505-518.

Parncutt, R. and McPherson, G.E. (eds.) (2002), *The Science and psychology of music performance: Creative strategies for teaching and learning*, OUP, Oxford.

Penel, A. and Drake, C. (1998), 'Sources of timing variations in music performance: A psychological segmentation model', *Journal of Psychological Research*, 61, 12 – 32.

Penel, A. and Drake, C. (1999), 'Seeking 'One' Explanation for expressive timing', *in* Yi, S.W. (ed.), *Music, Mind and Science*, Seoul, Seoul National University Press.

Peyser, H.F. (1925), '"Tristan", first hand', *Musical Quarterly*, xi, 418.

Poesio, M. and Vieira, R. (1998), 'A corpus-based investigation of definite description use', *Computational Linguistics*, 24, 183-216.

Posner, M.I. and Boise, S.J. (1971), 'Components of attention', *Psychological Review*, 78, 391-408.

Posner, M.I. and Klein, R.M. (1973), 'On the functions of consciousness', *in* Kornblum, S. (ed.), *Attention and Performance IV*, New York, Academic Press.

Povel, D.J. (1977), 'Temporal structure of performed music. Some preliminary observations', *Acta Psychologica*, 41, 309-320.

Reber, A.S. (1973), 'Locating clicks in sentences: Left, center, and right', *Perception and Psychophysics*, 13, 133-138.

Reimann (1903), *System der musikalischenm Rhythmik und Metrik*, Leipzig, Breitkopf und Härtel.

Reimann, H. (1884), *Musicalische Dynamik und Agogik. Lehrbuch der musikalischen Phrasirung auf Grund einer Revision der Lehre von der musikalischen Metrik und Rhythmik*, Hamburg.

Repp, B.H. (1990), 'Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists', *Journal of the Acoustical Society of America*, 88, 2, 622-641.

Repp, B.H. (1992), 'Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei"', *Journal of the Acoustical Society of America*, 92, 5, 2546-2568.

Repp, B.H. (1995), 'Expressive timing in Schumann's "Träumerei:" an analysis of performances by graduate student pianists.' *Journal of the Acoustical Society of America*, 98, 2413-2427.

Repp, B.H. (1999), 'A microcosm of musical expression: II. Quantitative analysis of pianists' dynamics in the initial measures of Chopin's Etude in E major', *Journal of the Acoustical Society of America*, 105, 1972-1988.

Réti, R. (1951), *The Thematic Process in Music*, New York.

Réti, R. (1958), *Tonality, Atonality, Pantonality: a Study of some Trends in Twentieth Century Music*, London and New York.

Riemann, H. (1884), *Musikalische Dynamik und Agogik: Lehrbuch der musikalischen Phrasirung auf Grund einer Revision der Lehre von der musikalischen Metrik und Rhythmik.*, Hamburg, Rather.

Rietveld, T. and Hout, R.v. (1993), *Statistical techniques for the study of language and language behaviour*, Berlin, Mouton de Gruyter.

Rink, J. (1995), 'The practice of performance'.

Rockstro, W.S., Dyson, G., Drabkin, W., Powers, H.S. and Rushton, J. (eds.), *'Cadence', Grove Music Online.*

Rockstro, W.S., Dyson, G., Drabkin, W., Powers, H.S. and Rushton, J. (1995), 'Cadence', *in* Macy, L. (ed.), *The New Grove Dictionary of Music and Musicians.*

Rosch, E. (1975), 'Cognitive representation of semantic categories', *Journal of Experimental Psychology: General*, 104, 192-232.

Rosen, C. (1997), 'Schubert's inflections of Classical form', *The Cambridge Companion to Schubert*, Cambridge, CUP.

Rosen, C. (2002), *Beethoven's Piano Sonatas: A short companion*, New Haven, Yale University Press.

Rosner, B. and Meyer, L.B. (1986), 'The perceptual roles of melodic, process, contour, and form', *Music Perception*, 4, 1-40.

Rothstein, W. (1981), 'Rhythm and the Theory of Structural Levels', Yale University.

Rothstein, W. (1988), 'Phrase rhythm in Chopin's nocturnes and mazurkas', *in* Samson, J. (ed.), *Chopin studies*, Cambridge, CUP.

Rothstein, W.N. (1989), *Phrase Rhythm in Tonal Music*, New York, Schirmer Books, A division of Macmillan Inc.

Rousseau, J.J. (1767), *Dictionnaire de Musique.*

Sadie, S. (1982), *The New Grove Mozart*, London, Macmillan.

Saffran, J.R., Loman, M.M. and Robertson, R. (2000), 'Infant Memory for Musical Experiences', *Cognition*, 77, B16–23.

Salzer, F. (ed.) (1987), *The Music Forum, Volume VI Part 1*, Columbia University Press, New York.

Samson, J. (1994), *The Cambridge Companion to Chopin*, Cambridge, CUP.

Samson, J. (1996), *Master Musicians: Chopin*, Oxford, OUP.

Schachter, C. (ed.) (1987), *Rhythm and Linear Analysis: Aspects of Meter*, Columbia University Press, New York.

Schaefer, R.S., Murre, J.M.J. and Bod, R. (2004), 'Limits to universality in segmentation of simple melodies', *in* Lipscomb, S.D., Ashley, R., Gjerdingen, R.O. and Webster, P. (eds.), *International Conference on Music Perception and Cognition (ICMPC8)*, Evanston, USA, Causal Productions.

Schenker, H. (1906, 1954), *Harmony*, Cambridge, MIT Press.

Schenker, H. (1906/1922/1935), *Neue musikalische Theorien und Phantasien*, Vienna.

Schenker, H. (1925, 1994), 'Abolish the Phrasing Slur', *The Masterwork in Music vol. 1 (Das Meisterwerk in der Musik, Munich, 1925)*, Cambridge, CUP.

Schenker, H. (1935, 1979), *Free Composition (Der freie Satz)*, New York, Longman.

Schmuckler, M. (1989), 'Expectation in music: Investigation of melodic and harmonic processes', *Music Perception*, 7, 109-150.

Schoenberg, A. (1960), 'The Orchestral Variations, Op. 31: A Radio talk, Frankfurt Radio, 1931', *The Score*, 28.

Schoenberg, A. (1975), *Style and Idea: Selected writings of Arnold Schoenberg*, Berkeley, Los Angeles.

Schulz, J.P.A. (1771), 'Performance', *in* Sulzer, J.G. (ed.), *Theorie der schönen Kunste (Theory of the fine arts) (1771-4).*

Seashore, C.E. (1938), *Psychology of Music*, New York, McGraw-Hill.

Sessions, R. (1950), *The musical experience of composer, performer and listener*, Princeton, Princeton University Press.

Shaffer, L.H. (1980), 'Analysing piano performance: A study of concert pianists', *in* Requin, G.E.S.J. (ed.), *Tutorials in motor behavior*, Amsterdam, North-Holland.

Shaffer, L.H. (1984), 'Timing in solo and duet piano performances', *Quarterly Journal of Experimental Psychology*, A36, 577-95.

Shaffer, L.H. (1985), 'The expressive component in musical performance', *Paper delivered at the Third International Conference on Event Perception and Action*, Uppsala, Sweden.

Shaffer, L.H. and Todd, N.P.M. (1987), 'The interpretative component in musical performance', *in* Gabrielsson, A. (ed.), *Action and Perception in Rhythm and Music*, Stockholm, Sweden, Royal Swedish Academy of Music.

Siegel, S. and Castellan, J. (1988), *Nonparametric Statistics for the Behavioural Sciences.*, McGraw-Hill Book Company.

Sloboda, J. (1983), 'The communication of musical metre in piano performance', *Quarterly Journal of Experimental Psychology*, A35, 377-96.

Sloboda, J. (1986), *The Musical Mind: The cognitive psychology of music*, Oxford, Clarendon Press.

Sloboda, J. and Deliège, I. (1996), *Musical Beginnings: Origins and development of musical competence*, Oxford, OUP.

Sloboda, J.A. and Gregory, A.H. (1980), 'The Psychological Reality of Musical Segments', *Canadian Journal of Psychology*, 34, 3, 274-280.

Snyder, B. (2000), *Music and Memory: An Introduction*, Massachusetts, The MIT Press.

Spiro, N. (2002), 'Combining Grammar-based and Memory-based models of Perception of Time signature and Phase', *in* Anagnostopoulou, C., Ferrand, M. and Smaill, A. (eds.), *International conference on music and artificial intelligence (ICMAI)*, Edinburgh.

Steedman, M. (1977), 'The perception of musical rhythm and metre', *Perception*, 6, 555-569.

Stein, D. and Spillman, R. (1996), *Poetry into song: Performance and Analysis of Lieder*, New York, Oxford, OUP.

Stewart, M. (1912), *Studies in phrasing and form*, London, Geo Barber.

Stobart, H. and Cross, I. (2000), 'The Andean Anacrusis? Rhythmic structure and perception in Easter songs of Northern Potosí, Bolivia', *British Journal of Ethnomusicology*, 9, 2, 63-94.

Stoffer, T., H. (1985), 'Representation of Phrase Structure in the Perception of Music', *Music Perception*, 3, 2, 191-220.

Street, A. (1989), 'Superior myths, dogmatic allegories: The resistance to musical unity', *Music Analysis*, 8, 1-2, 77-123.

Sundberg, J. (1988), 'Computer synthesis of music performance', *in* Sloboda, J.A. (ed.), *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, Oxford, Clarendon Press.

Sundberg, J. (2000), 'Grouping and differentiation. Two main principles in the performance of music', *in* Nakada, T. (ed.), *Integrated human brain science: Theory, method, application (music)*, Amsterdam, Elsevier.

Temperley, D. (2001), *The Cognition of Basic Musical Structures*, Cambridge, MA, MIT Press.

Temperley, D. and Bartlette, C. (2002), 'Parallelism as a Factor in Metrical Analysis', *Music Perception*, Vol. 20, no. 2, 117-149.

Tenney, J. and Polansky, L. (1980), 'Temporal Gestalt Perception in Music', *Journal of Music Theory*, 24, 205-241.

Thomson, W. (2001), 'Deductions Concerning Inductions of Tonality', *Music Perception*, 19, 1, 127-138.

Tiersot, J. (1895), 'Le chant du berger de Tannhäusser et celui de Tristen et Iseult', *Le ménestraek*.

Timmers, R. and Honing, H. (2002), 'On music performance, theories, measurement and diversity', *in* Belardinelli, M.A. (ed.), *Cognitive Processing (International Quarterly of Cognitive Sciences)*.

Titchener, E.B. (1909), *Lectures on the experimental psychology of the thought process*, New York, Macmillan.

Todd, N.P. (1985), 'A model of expressive timing in tonal music', *Music Perception*, 3, 33-58.

Todd, N.P.M. (1992), 'The dynamics of dynamics: A model of musical expression', *Journal of the Acoustical Society of America*, 91, 6, 3540-3550.

Todd, N.P.M. (1995), 'The kinematics of musical expression', *Journal of the Acoustical Society of America*, 97, 3, 1940 - 1949.

Tolman, E.C. (1948), 'Cognitive maps in mice and men', *The Psychological Review*, 55, 4, 189-208.

Tovey, D. (1944), *Musical Articles from the Encyclopaedia Britannica*, Oxford, OUP.

Truscott, H. (1963), 'Wagner's Tristan and the Twentieth Century', *The Music Review*, 24, 75 - 85.

Türk, D.G. (1789), *Klavierschule, oder, Anweisung zum Klavierspielen für Lehrer und Lernende*, Leipzig, Schwickert.

Vendler, H. (1997), *Poems, Poets, Poetry: An Introduction and Anthology*, Boston, Bedford Books.

Webern, A. (1963), *The Path to the New Music*, Bryn Mawr, Penn.L Thoedore Presser.

Wertheimer, M. (1924), 'Gestalt Theory, [Über Gestalttheorie], an address before the Kant Society, Berlin, 7th December 1924 ', *in* Ellis, W.D. (ed.), *Source Book of Gestalt Psychology, 1938*, New York, Harcourt, Brace and Co.

Westergaard, P. (1975), *An introduction to tonal theory*, New York, Norton.

Wittmann, M. and Pöppel, E. (1999), 'Time in the mind', *The MIT encyclopaedia of the cognitive sciences CD ROM*, MIT Press.

Zbikowski, L.M. (2002), *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*, Oxford, OUP.

Titles in the ILLC Dissertation Series:

ILLC DS-2001-01: **Maria Aloni**
Quantification under Conceptual Covers

ILLC DS-2001-02: **Alexander van den Bosch**
Rationality in Discovery - a study of Logic, Cognition, Computation and Neuropharmacology

ILLC DS-2001-03: **Erik de Haas**
Logics For OO Information Systems: a Semantic Study of Object Orientation from a Categorial Substructural Perspective

ILLC DS-2001-04: **Rosalie Iemhoff**
Provability Logic and Admissible Rules

ILLC DS-2001-05: **Eva Hoogland**
Definability and Interpolation: Model-theoretic investigations

ILLC DS-2001-06: **Ronald de Wolf**
Quantum Computing and Communication Complexity

ILLC DS-2001-07: **Katsumi Sasaki**
Logics and Provability

ILLC DS-2001-08: **Allard Tamminga**
Belief Dynamics. (Epistemo)logical Investigations

ILLC DS-2001-09: **Gwen Kerdiles**
Saying It with Pictures: a Logical Landscape of Conceptual Graphs

ILLC DS-2001-10: **Marc Pauly**
Logic for Social Software

ILLC DS-2002-01: **Nikos Massios**
Decision-Theoretic Robotic Surveillance

ILLC DS-2002-02: **Marco Aiello**
Spatial Reasoning: Theory and Practice

ILLC DS-2002-03: **Yuri Engelhardt**
The Language of Graphics

ILLC DS-2002-04: **Willem Klaas van Dam**
On Quantum Computation Theory

ILLC DS-2002-05: **Rosella Gennari**
Mapping Inferences: Constraint Propagation and Diamond Satisfaction

ILLC DS-2002-06: **Ivar Vermeulen**
A Logical Approach to Competition in Industries

ILLC DS-2003-01: **Barteld Kooi**
Knowledge, chance, and change

ILLC DS-2003-02: **Elisabeth Catherine Brouwer**
Imagining Metaphors: Cognitive Representation in Interpretation and Understanding

ILLC DS-2003-03: **Juan Heguiabehere**
Building Logic Toolboxes

ILLC DS-2003-04: **Christof Monz**
From Document Retrieval to Question Answering

ILLC DS-2004-01: **Hein Philipp Röhrig**
Quantum Query Complexity and Distributed Computing

ILLC DS-2004-02: **Sebastian Brand**
Rule-based Constraint Propagation: Theory and Applications

ILLC DS-2004-03: **Boudewijn de Bruin**
Explaining Games. On the Logic of Game Theoretic Explanations

ILLC DS-2005-01: **Balder David ten Cate**
Model theory for extended modal languages

ILLC DS-2005-02: **Willem-Jan van Hoeve**
Operations Research Techniques in Constraint Programming

ILLC DS-2005-03: **Rosja Mastop**
What can you do? Imperative mood in Semantic Theory

ILLC DS-2005-04: **Anna Pilatova**
A User's Guide to Proper names: Their Pragmatics and Semanics

ILLC DS-2005-05: **Sieuwert van Otterloo**
A Strategic Analysis of Multi-agent Protocols

ILLC DS-2006-01: **Troy Lee**
Kolmogorov complexity and formula size lower bounds

ILLC DS-2006-02: **Nick Bezhanishvili**
Lattices of intermediate and cylindric modal logics

ILLC DS-2006-03: **Clemens Kupke**
Finitary coalgebraic logics

ILLC DS-2006-04: **Robert Špalek**
Quantum Algorithms, Lower Bounds, and Time-Space Tradeoffs

ILLC DS-2006-05: **Aline Honingh**
The Origin and Well-Formedness of Tonal Pitch Structures

ILLC DS-2006-06: **Merlijn Sevenster**
Branches of imperfect information: logic, games, and computation

ILLC DS-2006-07: **Marie Nilsenova**
Rises and Falls. Studies in the Semantics and Pragmatics of Intonation

ILLC DS-2006-08: **Darko Sarenac**
Products of Topological Modal Logics

ILLC DS-2007-01: **Rudi Cilibrasi**
Statistical Inference Through Data Compression

ILLC DS-2007-02: **Neta Spiro**
What contributes to the perception of musical phrases in western classical music?