COMPUTER AIDED STATISTICAL ANALYSIS OF MOTIVE USE AND COMPOSITIONAL IDIOM

Peter Anthony Smith

A thesis submitted in partial fulfilment of
requirements for the degree of
Master of Music (Performance)
Sydney Conservatorium of Music
University of Sydney
2012

I declare that the research presented here is my own original work and has not been submitted to any other institution for the award of a degree.

Signed: ....

Date: ....

Abstract

This thesis discusses the creation of a means of pitch-based data representation which allows automated logging and analysis of melodic motivic material. This system also allows analysis of a number of attributes of a composition which are not readily apparent to human analysis. By using a numerical data format which treats motivically related material as equivalent, groups of tonally equivalent intervals (n-tuples) can be logged and have statistical procedures carried out on them. This thesis looks at four applications of this approach: measuring the most commonly occurring motivic material; creating a transition matrix showing probabilities of movement between intervals; measuring the extent of disjunct or conjunct writing; and measuring concentration of motivic writing (the extent to which motives are reused). Following the discussion of the data representation system, a set of expositions taken from the piano sonatas of Haydn, Mozart, and Clementi are converted to this method of data representation, and results are collected for the above four applications. The implications of the results of this analysis are discussed, and further potential applications of the system are explored.

Acknowledgements

I would like to thank Jen Jen Chung and Matthew Woolhouse for their assistance in the early stages of this thesis. Their critical responses to ideas helped to give this research its present focus.

I owe a debt of gratitude to my parents, Kingsford and Linda, for supporting me throughout my musical education. My brother, Philip, has been an invaluable expert in regards to programming, mathematics, and academia in general.

To my partner, Erin Moss, I owe thanks for providing considerable assistance with much of the data entry involved in this research, and for being a constant source of encouragement and patience.

Perhaps most of all I thank my supervisor, David Larkin, who has been an incredible inspiration to me for as long as I have known him. His critical mind and attention to detail have helped to shape and improve this thesis constantly.

Lastly, I thank the Sydney Conservatorium of Music for their facilities and the funding to pursue this research.

Contents

1 - Introduction

1.1 - Existing Related Research

Human analysis of music, whatever form it takes, is a highly subjective process guided by complex rules and aesthetic inferences. In addition to requiring a thorough knowledge of stylistic idioms, numerous exceptional cases require the formation of intelligent personal judgements. Given the inherent complexity in defining the very *task* of musical analysis, many computer-science researchers have found that the process of automating musical analysis offers a wide variety of fascinating challenges.

Approaches towards automating musical analysis with computers were first proposed in the 1960s, where an initial wave of research into computer-based Natural Language Processing (NLP) was taking place. A significant early paper by Winograd (1968) discusses the application of rules relating grammatical structures to music analysis, and looks at issues of information representation (an issue which continues to be a contested area in all research in the field). After the 'AI Winter' — a period from the late 1960s to early 1980s during which funding for artificial-intelligence research was difficult to obtain due to a lack of grants from the American government[1] — numerous other theorists worked on the problems of computer-based analysis from different perspectives. Several writers attempted to consolidate existing research while looking forward: Meehan (1980) draws heavily from early grammatical research, while Roads (1980) expands the scope of computer analysis discourse to include algorithmic composition. Such areas received greater attention in the years to follow, becoming a significant branch of 'AI Creativity'. While not directly related to computer analysis, Lerdahl and Jackendoff's influential book *A Generative Theory of Tonal Music* (published in 1983) outlined an algorithmic means of parsing musical structures of all lengths. This study has been cited by almost every subsequent paper discussing automated analysis.

Many papers from the 1980s do not discuss the possibility of an analytical program making complex aesthetic judgements (such as the ability to make decisions about ambiguous structures, or partitioning areas of different tonalities) and are more concerned with creating

1 While research in this field is now occurs globally, most early artificial-intelligence research originated in the United States; hence the unique importance of American research funding.

tools for human theorists to use. Brinkman (1986) discusses the use of the DARMS notational format to represent scores for computer analysis, and Smoliar (1980) looks at the preparation of Schenkerian analyses using transformation programs. Schenkerian analysis has been a salient problem in AI analysis, and many researchers have approached it in the last decade — examples include Kirlin's (2009) work to produce foreground reductions based on algorithms by Pardo and Birmingham (2002), and Marsden's (2007) software for gradually reducing structural complexity.

Intimately tied up with the area of musical analysis is computer-based composition. There have been innumerable approaches to this topic, and many of these studies are not limited merely to traditional tonal systems. While a discussion of the entirety of the field of AI composition is outside the scope of this literature review, numerous papers make links to computer analysis by citing statistical systems or algorithmic construction techniques and are worth mentioning. In his article *Computer Improvisation* (1980), Fry creates a robust system for the generation of improvised jazz music which can have many factors modified at both the global and local scale. In one of the few papers attempting to examine human response to computer-generated music, Brown (2004) creates melodies using algorithmic and genetic systems, and analyses their aesthetic quality. Much research has been undertaken examining chorale melody generation, as it is a field where compositional principles are easily examined by human judges. Pearce (2005) uses a variety of statistical approaches to generate stylistically appropriate chorale melodies, and Pearce and Wiggins (2007) use perceptual heuristics based on the feedback of expert judges to improve the quality of chorale melodies generated by a Markov chain system. One of the most interesting bodies of work related to algorithmic music generation has been undertaken by David Cope. Cope has put forward a program which is capable of analysing atonal voice leading parts and creating rule sets from this input to produce new polyphonic music (2002). In Cope's *Virtual Music: Computer Synthesis of Musical Style* (2001), he details the workings of the *Experiments in Musical Intelligence* program — a piece of software using various perceptual heuristics which is capable of mimicry of arbitrary stylistic features at both a surface and deep structural level. More recently, Cope wrote a paper which outlines how a compositional AI can learn rules from an arbitrary corpus, and quickly overtake human proficiency in basic contrapuntal writing (2004). In the last decade, papers have started to acknowledge that any program which imitates human compositional processes will necessarily feature complex algorithmic processes, even if it is trained on a corpus of works. Dubnov et al. (2003) discuss how compositional programs have rapidly increased in complexity and success since the 1980s, and they put forward their own programs which are capable of extremely flexible and realistic composition in a variety of genres.

From the late 1990s to the present day there has been a great amount of research in divergent areas related to computer-based musical analysis. This research often justifies its relevance in the relatively new field of Music Information Retrieval (MIR). Since analytical AI research is no longer new, papers have either begun to approach more esoteric subject matter, solve problems which have historically been more difficult, or use more advanced mathematical and statistical tool sets. Pachet (1997) uses hierarchical element trees combined with numerous rules related to jazz composition to create a program capable of deducing musical structure even when such structures are obscured by extraneous elements. This has the potential to create a quantifiable computational metric of any music structure, and connects much more deeply with the notion of large scale structures than studies had previously. Hörnel and Menzel created a system they titled HARMONET, which uses trained neural networks to harmonise melodies in the style of an arbitrary composer. This concept was then expanded to create variations of melodies, using a system called MELONET. While it could be argued that functions such as this could be created using rule systems (which have been present since much earlier), works based on randomisation and rules are generally felt to be less aesthetically pleasing than works created using statistically trained neural net systems (Hörnel & Menzel, 1998). In addition, while their study was primarily focused on composition, analysis of the weightings of neural net nodes reveals idiomatic patterns in composers' outputs. This is of great interest to analysts seeking to understand the concept of 'style'. Several researchers have turned their attention to testing phrase boundary algorithms outlined in Lerdahl and Jackendoff's *A Generative Theory of Tonal Music*, and testing other statistical methods of perceptual grouping. Pearce and Wiggins have contributed heavily in this field; in a paper from 2008 (with Müllensiefen), they compared statistical and rule based methods of melodic segmentation; in 2006, they proposed their own method of boundary prediction based around unpredictability of subsequent melodic events (linking to the Gestalt psychology research mentioned on page 6). Thom et al. (2002) compared the phrase structures generated by several researchers' algorithms to phrase structures intuitively created by professional musicians. There are many more novel and new approaches to old problems that could be listed in addition to these, and the number is growing. To name a short few: Raphael (2003) uses probabilistic graphs in determining correct analysis of harmony; Abdallah et al. (2005)

uses Bayesian models to elicit musical structures; Manaris et al. (2007) look at how previous mathematical methods of analysis can be united with an approach based around repertoire learning; and Whorley et al. (2004) uses element cross-matching predictive systems operating on numerous musical criteria in the study of four-part harmony.

Despite the many attempts to devise reliable means of assessing harmony and (stemming from this) larger-scale structure, comparatively little research has been undertaken into any other forms of analysis. At the time of writing, there has been no academic writing regarding Neo-Riemannian AI analysis tools, and work related to semiotic or narratological analysis is (at least for now) quite impractical. While some research makes use of set theory, the mathematical reductions required to transfer sonorities into interval class vectors is trivial. The larger issue in analysis of post-tonal works is probably deducing points of structural division, and this is a problem which invites analogous approaches to those employed in tonal music.

Artificial-intelligence based motivic analysis is a thriving field of research which is separate to the previously discussed tonal and structural analytical systems. There are two main researchers who are linked to almost every significant paper in this area so far — Emilios Cambouropoulos in Austria, and Olivier Lartillot in Finland. Cambouropoulos has done research into algorithms usable for determining similarity and relatedness of musical material (Cambouropoulos & Widmer, 2000), and hence produced an interesting structural analysis of Schumann's *Träumerei* (Cambouropoulos, 2000). He acknowledges the significant overlap that motivic analysis has with perception and psychology, and so avoids making definite labels and classes for motivic relations — they are simply organised in priority based on context. Lartillot has produced algorithms capable of finding patterns in music, and solves intractability issues by a variety of heuristics to avoid combinatorial explosion (Lartillot, 2005b, 2005c). In acknowledging the significant extent to which perception influences motivic analysis, Lartillot also tries to recreate human cognitive means of 'redundancy suppression' — a series of heuristic algorithms which exclude insignificant motivic patterns from analysis (Lartillot, 2005a).

In addition to traditional types of analysis, the second half of the twentieth century saw the integration of information theory into traditional musicology. Information theory is a field of mathematics dealing with the transferral of information in signal form. It is a field which utilises statistical analysis to understand information redundancy and entropy[2] in any piece of data. Applying its processes to music using statistical logging of musical events yields some surprising insights and reinforces inferences from traditional analysis. Meyer (1957) was among the first to discuss the possibilities for information theory analysis to quantify stylistic idioms. Following this, Youngblood (1958) applies information theory analysis to melodies by Schumann, Mendelssohn and Schubert and discusses the significance of his statistical results. Information theory has been proposed as an automated means of separating musical styles and genres: research from the 1950s and 1960s suggests that the difference in idiolects between composers might be found in signal information analysis. In addition to this application, however, it has also been used as an analytical tool in its own right. Hiller & Bean (1966) produced graph-based analyses of four contrasting sonata expositions which can be interpreted to give information about the intensity or chromaticism of different pieces of music (or different sections within a single composition). Hiller & Ramon (1967) then followed this study with an in-depth analysis of Webern's *Symphonie* Op. 21, comparing traditional structural analysis with an information theory analysis. While information theory as a field does not require computers, the statistical logging and processing which it requires are far more easily carried out with the aid of automated programs.

While information theory itself relies on statistics, there have been numerous studies which can be seen as purely statistical. In addition to the studies already mentioned which make use of graphical models using statistics-based learning, some researchers have attempted to quantify features of music based on event frequencies: Toiviainen & Eerola (2001) used a self-organising map on a large collection of folk songs to discern regional differences, while Manaris et al. (2005) established how composers can be identified by the extent to which their output has features following a Zipf distribution (Zipf distributions and their relevance to this study will be covered in sections 2.4.1 and 3.2.1). Most statistics-based musical analysis is conducted on audio signals with application in information retrieval, such as Beran & Mazzola's study (1999) which used statistics to draw conclusions regarding the musical structure of a composition from a collection of recordings. There have also been attempts to create systems allowing clear quantification of stylistic features. A thesis by Bellman (2011) uses detailed information about musical score events in combination with statistical pattern recognition algorithms to form clear metrics when analysing musical genres. As a result of this, his system proved capable of predicting the composer of an unknown work (comparing it

2 Entropy, as the term is used in information theory, is analogous to the amount of possible combinations of stored data in any given signal. In musical information theory analyses, it is generally seen as a measure of complexity.

to data from a representative corpus) with up to 75% accuracy, even when operating on works of significant perceptual similarities (such as, for instance, separating some works of Beethoven and Mozart).

Statistical analysis of scores has also often been used to test hypotheses on the nature of human musical perception. Vos (1989) carried out a study on melodic intervals in Western music and found that descent by scale was a generalised contour. This was then experimentally tested against subjects' perceptual inferences. In the same year, Eugene Narmour published a formal outline of the 'implication-realization' model of melodic prediction, which argued that there are certain sequences of intervallic direction and magnitude which humans naturally find predictable and satisfying, following principles similar to those from Gestalt theories of visual analysis. This model has been experimentally tested and verified in a number of psychological and computational studies. Schellenberg (1996) showed that the 'implication-realization' model explained melodic prediction tendencies in test subjects regardless of the ethnicity of their upbringing or their level of musical education. This was further proved in research using melody creativity by Thompson et al. (1997). Thompson and Stainton (1998) proved the accuracy of both Narmour's closural and implication theories in a large corpus of folk songs, while Pearce and Wiggins (2004b) used a statistical analysis of folk songs, ballads, and chorale melodies to show that Narmour's supposed instinctive principles may simply be the result of exposure to a wide corpus of music which embodies Narmour's contour tendencies. Narmour also theorised that melodic expectation arises not only from intrinsic human nature, but through long term cultural exposure (extraopus style) and short term knowledge about a piece of music as it is being listened to (intraopus style). Most studies referencing the 'implication-realization' theory focus on the innate associations it outlines, but studies such as Thompson et al. (2000) have also psychologically tested the validity of intraopus information. Since the 'implication-realization' model is very clearly defined, it is straightforward to produce annotated automated analyses of melodies, showing their structure in terms of Narmour's intervallic transitions. Grachten et al. (2005) did just this, creating a system which was very capable of demonstrating structural similarities in melodic construction.

In addition to studies which provide evidence for Narmour's theory, there also exists a significant body of work devoted to inductively forming broad hypotheses regarding general melodic contour. In *Sweet Anticipation* (2006), David Huron's book discussing musical expectancy, an analysis of folk song corpora produced surprisingly consistent contour data, even though most of this information is already understood intuitively by any musical professional. While results differed slightly with folk song corpora from different regions, Huron shows that there are several unanimous results:

- Stepwise motion predominates, accounting for the largest proportion of intervallic motion in all corpora by far.
- There is a general pattern that ascents by leap are followed by descents by step (interestingly, Huron also makes the point in a paper from 2000 that this is largely due to the effect of vocal ranges on folk song composition).
- Stepwise motion tends to precede stepwise motion.

Many of these consistent observations are supported by Narmour's theory. These findings were further reinforced in David Temperley's *Music and Probability* (2007). An analysis of the Essen folk song collection (comprising 6252 folk songs of mostly Germanic and European origin) revealed that roughly 40% of material was accounted for by descending step and unison intervals, and slightly less than 10% was accounted for by ascending step. Temperley's research then proceeds to attempt quantification of subtle musical phenomena (such as ambiguity, tonalness, and tension) and to further the argument of Pearce and Wiggins that experience is the most significant factor in shaping perceptual expectation. In Huron (1996), another exhaustive study of the Essen folk song collection, the intuitive hypothesis that Western music largely possesses an arch shaped phrase contour is supported through statistical analysis. This paper also shows a relatively early generalisation of intervallic movement which is in concordance with the findings above.

1.2 - Research Aims

The research presented in this thesis is related to existing work in the field of motivic analysis, while also drawing from research which attempts to quantify compositional idiolects through statistical analysis. Like all of the work presented by Lartillot and Cambouropoulos, and much of the existing statistics-based work, this thesis focuses solely on analysis of monodic structures. When the source being examined originally features a homophonic texture, the melodic line is first extracted to form a monodic reduction. The systems used in this thesis do also not take into account any rhythms or rhythmic relationships. The reasons for this are discussed in section 2.1.

It is not the intention of this thesis to propose alternate methods of motivic analysis to those put forward by Cambouropoulos and Lartillot. Previous research in that field has attempted to use motivic analysis to propose structural divisions within a piece, or to test the extent of relationships between different pieces of musical material. Recent studies have proposed an algorithmic means of pattern recognition which models human perception. Areas of research such as these, while wholly legitimate, are outside the scope of this study. Additionally, this thesis does not intend to incorporate information theory analysis, as this is a field which has already been thoroughly pursued by other researchers. However, this research does draw upon some of the existing statistical work done by researchers such as Youngblood, Leon & Hutchinson (1983), and Hiller, as it shares their concern with the quantification of the concept of 'style'. This research employs statistical analysis procedures on patterns which have significance as small musical structures, examining their re-use and what this data may imply about the concept of musical style.

Unlike the approach adopted by Bellmann and many others, this research relies on the analysis of scores without making use of tonality or scale degrees, in order to keep the motivic information it generates in a very practical generic form. This thesis details the construction of a data format into which monodic music is converted. In the process of conversion, motives[3] which are diatonically similar (but may be chromatically different) are treated as equal entities. This data is scanned by a program, and tables of recurrent motivic sequences are created. These sequences, referred to as *n*-tuples (intervallic sequences featuring *n* number of notes) form almost the entirety of the raw data which is analysed. Statistical operations are then carried out on these sequences which quantify significant aspects of composition which are not easily humanly assessed. These statistical assessments allow comparison of melodic idioms between corpora of works — section 2.4 details four ways in which the data can be compared.

A main outcome of this research will be to test the concordance of results with the verified hypotheses put forward by Cope, Huron, Temperley, Narmour, Pearce and Wiggins, and others. Comparison will be drawn between the results of this research operating on a classical-piano corpora, to those bodies of work operating on folk song corpora.

3 This thesis uses the term 'motive' as it is understood in texts by Schoenberg: a small amount of musical material which gains significance by its re-use and development. Note, however, that the term is also used here to describe n-tuple intervallic groupings which may occur as little as once, and have no greater significance within a work. The term 'motive' is used despite any perceptual significance that may connote, because despite many *n*-tuples having no perceptual significance, they may still have relevance as repeated material.

The corpora of works used in this thesis come from the piano sonatas of Mozart, Haydn, and Clementi. The rationale behind this choice of repertoire is discussed in section 3.1. From the data gathered from these corpora, generalised observations regarding melodic idiom are formed, and a means of discriminating between the attributes of the melodies of these composers is stated. From these results, the idiolects of individual composers is quantified, and a means of comparing their motivic vocabulary is given.

2 - Methodology

While the works which this thesis intends to examine are all essentially homophonic (featuring an admixture of polyphonic elements), statistical operations in this study are all carried out on purely monodic lines. This is done with the intention of isolating features of melodic idiom, rather than any other features related to texture or harmony. The first step in subjecting these works to statistical tests is the reduction of their homophonic textures to single melodic lines.

2.1 - Melodic Line Reduction

It is the intention of this thesis to examine melodic lines without imposing any perceptual or algorithmic heuristics to limit the material which is analysed. In order to avoid having to make subjective decisions about the textural identity of voices in a piece of music, works being analysed by the system outlined by this study should feature a constant melodic voice in a continuous register. The texture of many works from the classical period features a clear distinction between melody and accompaniment functions. While there are obviously numerous exceptions, it is widely recognised that most works of the classical era (in opposition to earlier periods) tend to feature a clear melody sitting above other accompanying parts (Burkholder, Grout, & Palisca, 2010, pp. 300-301). All of the works chosen for analysis in section 3 generally fit this description, but there are still several areas where the reduction of the music to a single melodic line requires subjective decision making, and at times, a concession to the local polyphony. Some of these examples are discussed in section 3.1.1, and an exhaustive list of all the issues arising in melodic reduction in this experiment is given in the appendix.

It must also be acknowledged that the task of identifying which of a set of voices is the melodic line is non-trivial. This is an issue that has been grappled with several times in music information retrieval and automated analysis (León, Rizo, & Iñesta, 2007; Madsen & Widmer, 2007b). One identifying criteria put forward by Madsen & Widmer et al. (2007a) is that of entropy (as the term is understood in the field of information technology). Under this system of identification, the more information that is required to express all the detail of a musical line, the more likely that line is to be melodic. This observation is probably valid for most classical era music (and many other genres) but issues still arise in the interchange of melody between different voices in a musical texture. Without the prior formation of partition points where it is known that the melody changes voices, this definition is still problematic for automated analysis. This research makes no attempt to establish strict guidelines for identifying melodic lines, and instead relies on a general educated listener's understanding of what constitutes a melody.

2.1.1 - On Rhythmic Significance

As stated in the research aims, this project ignores all rhythms in the process of melodic reduction and data representation. This does not betoken a lack of recognition of rhythmic importance in motivic analysis, but rather an acknowledgement of the many problems facing those who would attempt the incorporation of rhythmic data.

It stands to reason that rhythms should play a part in motivic analysis because rhythms are inherently motivic (and vice versa). Motives contain a rhythmic element, and classical works (such as the ones under analysis in this thesis) tend to re-use and develop rhythmic material heavily (Schoenberg, 1967, pp. 8-15). As a small example, consider figure 1, the famous opening of Mozart's *Serenade No. 13 in G:*

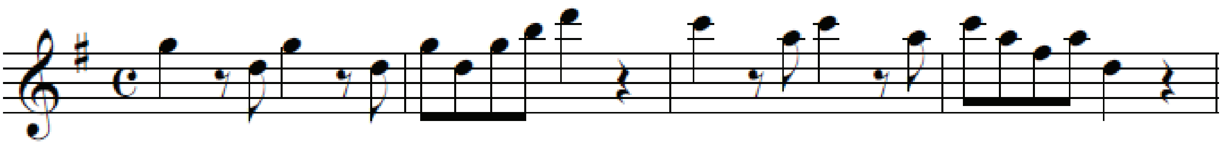


Fig. 1

The motivic relationship between the antecedant and consequent phrase in these four bars is reinforced by both pitch and rhythm. In terms of pitch, the consequent phrase is a non-exact inversion of the shape of the antecedant, and both move entirely within arpeggios (outlining the tonic in the antecedant, and the dominant in the consequent). However, the strongest link between these two phrases is definitely rhythmic — they are rhythmically identical. Perceptually, this feature is obvious, even to non-musicians. Even though the intervallic patterns used in each phrase are different, listeners are able to identify a motivic link between the phrases because of their rhythmic congruence.

If this research was attempting to analyse the relationship between motives, rhythmic relationships would obviously need to be taken into account. But in order to limit ‘false-positive’ automated tallies of material which aren't motivically significant, some rhythmic heuristic rules would need to be introduced. Potentially, only the motives which occur on metrically strong beats could be tallied, but this would discount anacruses, or rhythmic motives which are altered with anacruses. The rhythmic fragmentation of motives (as well as diminution or augmentation) is also difficult to account for without introducing other perceptual rules. The frequent ornamentation of classical motives with extra rhythmic embellishments adds to the difficulty of formulating a program to define ‘significant’ motives. Additionally, due to the open-ended nature of rhythmic interpretation for many ornaments, more issues arise if the rhythmic component of ornamentation is treated motivically.

Since a broad set of rules which could discern motivic significance based on rhythmic location are difficult to implement, this study ignores rhythm based analyses and only looks at a contiguous series of melodic pitches. Analysis of this sequence of pitches intends not to discern motivic relationships, but to form more general conclusions about the nature of melodic writing in a work (or corpus of works). While there will be numerous entries appearing in the motive-incidence tables which would not be considered motivic (for instance, groups of notes split between two phrases, or a portion of a theme which is rhythmically displaced) the greater number of incidences of the most significant pitch patterns will still rise to the top of any frequency tables.

Also, it should be noted that the size of motives being examined in this study is very small — only up to six pitches. This is enough size to identify common embellishing figures, short melodic motives, and common shapes in general melodic writing. It is not enough to examine most full themes or phrases. It is intended that through analysis of melodies from a corpus of works, the most common motivic pitch patterns will emerge. These short patterns can appear in any rhythm, and through studying them the pitch based idiolects of a composer can be identified more easily without considering metrical relationships.

2.2 - Data Representations

Once the work being analysed has been reduced to a single melody, and considering that rhythms do not factor into the data representation, the next step is to convert the melody into a useful format for motivic analysis. The primary aim of this data format is to render any contour motion which is motivically similar as identical, despite potential intervallic incongruence. This is done through the use of a series of numbers representing not specific pitches, but ‘staff position classes’ — a number which can stand for any note in a certain position on the staff. For instance, using scientific pitch notation, the notes C4, C♭4, and C♯4 are all the same staff position class — they all appear in the position of middle C on the staff, regardless of further chromatic designation. Double sharps and double flats also fall into the same staff class as the letter they are appended to. By using this designation, motivic relationships are maintained even when transpositions are not exact. For instance, the first three notes of C major in sequence are recognised as similar to the first three notes of C minor in sequence, because E and E♭ are the same staff position class.

These staff position classes have been given numerical designations. Labelling begins with 1.5 for C0 (and all possible chromatic inflections of this pitch which appear in that staff position) and rises by one for each ascending letter. The .5 decimal designation is used to assist in expressing some chromatic passages, and will be explained shortly. Continuing this system of designation yields the following table:

C0	1.5	**D0**	2.5	**E0**	3.5	**F0**	4.5	**G0**	5.5	**A0**	6.5	**B0**	7.5
C1	8.5	**D1**	9.5	**E1**	10.5	**F1**	11.5	**G1**	12.5	**A1**	13.5	**B1**	14.5
C2	15.5	**D2**	16.5	**E2**	17.5	**F2**	18.5	**G2**	19.5	**A2**	20.5	**B2**	21.5
C3	22.5	**D3**	23.5	**E3**	24.5	**F3**	25.5	**G3**	26.5	**A3**	27.5	**B3**	28.5
C4	29.5	**D4**	30.5	**E4**	31.5	**F4**	32.5	**G4**	33.5	**A4**	34.5	**B4**	35.5
C5	36.5	**D5**	37.5	**E5**	38.5	**F5**	39.5	**G5**	40.5	**A5**	41.5	**B5**	42.5
C6	43.5	**D6**	44.5	**E6**	45.5	**F6**	46.5	**G6**	47.5	**A6**	48.5	**B6**	49.5

This system of designation allows considerable flexibility for identifying motivic similarity not just diatonically, but also where a motive can be chromatically altered — consider figure 2, a short passage in C major:

Fig. 2

In each bar, there is a pattern which is repeated, diatonically transposed up a step. This pattern is ascending second, ascending second, descending third. Chromatically, in terms of semitone movement, each bar contains a different pattern, but the motivic relationship between each is obvious. Expressed in staff position classes, the passage becomes 29.5, 30.5, 31.5, 29.5, 30.5, 31.5, 32.5, 30.5, 31.5, 32.5, 33.5, 31.5. In this format, the +1, +1, -2 pattern of the passage is clear, and this pattern of movement is the same regardless of the key in which it is presented.[4]

4 Note that all references to intervallic movement in this thesis using positive and negative integers are not using the traditional musical notion of scale degrees, but of steps between notes. For example, using this notation, an ascending second is +1, while a descending fifth is -4.

In baroque, classical, or early romantic music, chromaticism most often arises through long and short term modulations and tonicisations, and through melodic embellishment (such as chromatic neighbour notes or chromatic passing tones). In this system of reduction, some chromatic melodic embellishment must necessarily find a means of representation, because not doing so would produce an extremely innaccurate depiction of the pitches involved. However, most chromaticism related to tonicisation is ignored to preserve motivic congruence. Figure 3 shows a short passage in which chromaticism is ignored:

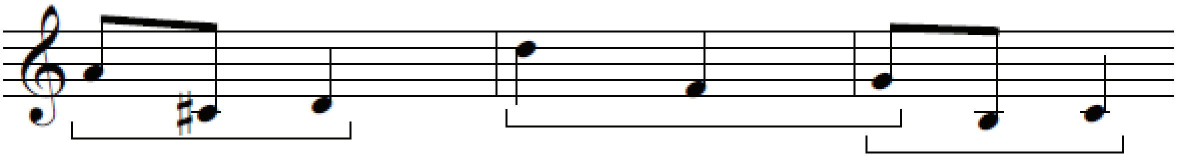


Fig. 3

In this passage, three occurrences of a motivically similar set of pitches have been shown with braces. Despite the presence of chromaticism, the motivic congruence between these three incidences is clear. The C♯ in the first bar is a non-diatonic note in C major, but can be dealt with by assigning it the same staff position class as the C♮ in the third bar. This reduces to the following numeric sequence: 34.5, 29.5, 30.5, 37.5, 32.5, 33.5, 28.5, 29.5. By representing the pitches this way, all three motivic occurrences have the same identity: -5, +1. This is why chromatic notes are generally ignored in the melodic reduction — to recognise incidences of repeated motives that can either be intervallically different within a single key, intervallically the same but in different keys, or slightly intervallically different and in different keys.

Some passages do however require chromaticism to be acknowledged in the reduction. This is most often seen in passages featuring ascending or descending chromatic scalic movement. Figure 4.1 and 4.2 are two passages featuring chromaticism.

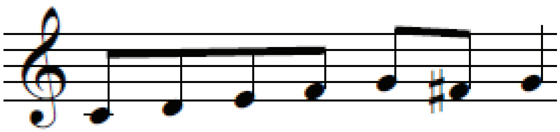


Fig. 4.1

Fig. 4.2

In figure 4.1, the F♯ arises as chromatic neighbour note to G. If the three note figure of the last three pitches is transposed onto an F or a C, it is no longer chromatic in C major. To preserve the relationships between these transposed motives, the F♯ is treated as any other staff position class on F, making figure 4.1 represented by 29.5, 30.5, 31.5, 32.5, 33.5, 32.5, 33.5. Some information is lost by this representation though — using that numeric sequence, the passage is considered the same as if it contained no chromatic note at all.

Reduction of chromatic elements into this data format is not always a straight-forward process, and subjective analytical decisions must sometimes be made. If, in figure 4.1, the last four notes, and not just the last three notes, are deemed to be motivically significant by their continued occurrence, then it could be argued that attempt should be made to better represent instances with that melodic contour in the motive frequency tables. In cases such as these, 0.25 can be added to a staff position class to represent a sharpened pitch, and 0.25 can be subtracted to represent a flattened pitch. By doing this, the non-decimal portion of a staff position class will always represent the underlying position of that note on the staff. Representing the data this way, the last four notes of figure 4.1 are 32.5, 33.5, 32.75, 33.5. The non-decimal portion of the F and F♯ (32) represents that position on the staff, and the motive is now uniquely represented.

Repeated chromatic passages where that kind of four-note figure re-occurs enough to be considered motivic are generally rare in the repertoire used experimentally in this thesis. It is far more common that the passages which require chromatic expression use passing chromatic notes, such as in figure 4.2. Here, without taking the passing F♯ into account, the reduction would be 29.5, 30.5, 31.5, 32.5, 32.5, 33.5, 29.5. This implies that the fourth pitch in this passage is immediately repeated, which is not the case. To better represent this, 0.25 is added to the fifth note in the passage, making 29.5, 30.5, 31.5, 32.5, 32.75, 33.5, 29.5. This is a much better representation of this passage, and any future occurrences of the same passage will share this motivic identity. Most examples of chromaticism in the repertoire analysed in this thesis feature chromatic scalic passages such as this one.

2.2.1 - Representation Limitations

This manner of reduction and representation is most effective on entirely diatonic music, and as music becomes increasingly chromatic, more issues in its data format emerge. While it is still highly functional for most classical melodies (hence the focus of this thesis), there are occasional problems with representation of chromatic passages. Consider the examples shown in figures 5.1, 5.2, and 5.3:

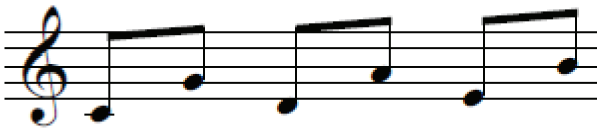


Fig. 5.1

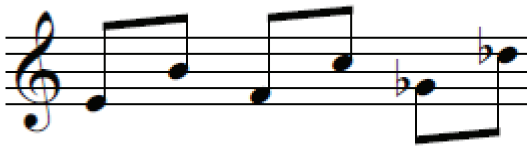


Fig. 5.2

Fig. 5.3

Figure 5.1 is simple and entirely tonal. It is easily represented numerically as 29.5, 33.5, 30.5, 34.5, 31.5, 35.5 or intervallically as +4, -3, +4, -3, +4. Similarly, under the systems outlined in the previous paragraphs, the passage in figure 5.2 should represent the same series of intervals when reduced to this data format (since the G♭ becomes 33.5 like all other notes built on G). However, figure 5.3 is intervallically identical to 5.2, yet can not be represented by the same intervals. Without making use of chromaticism on the C♯ and G♯, the passage can not logically be represented, because 29.5, 33.5, 29.5, 33.5, 30.5, 34.5 implies the first pitches are simply repeated. Considering that this data format intends to make similar motivic passages identical for purposes of analysis, it seems strange that two passages which are perceptually (ignoring transposition) identical are numerically different.

The justification of this is that while the passages in figures 5.2 and 5.3 are aurally extremely similar, as motivic constructs on a page they are significantly different. 5.2 features an ascending scale in fifths, while 5.3 is written with a clear chromatic passing note. Issues such as this arise due to the asymmetry of the tonal system and the non-identity of enharmony. Situations which are this contentious are relatively rare, however, and this data representation system is adequate for the expression of most motivic constructs in classical music.

Before concluding this section, a brief discussion must be made of the relevance of ‘signatures’, as defined in David Cope's *Virtual Music.* Cope defines signatures as “contiguous note patterns which recur in two or more works of a single composer and therefore indicate aspects of that composer's musical style”. Signatures are often perceptually similar figures which may appear on the surface to be quite different. Figure 6 shows two perceptually related examples from the piano sonatas of Mozart, given as an example by Cope. Example *a* is from K. 279, movement 2, and example *b* is from K. 280, movement 2.

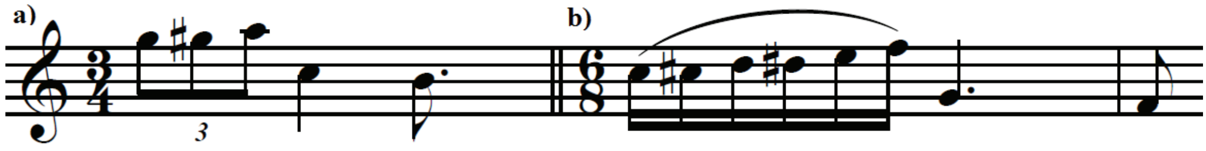


Fig. 6

While the number of notes, intervals, and rhythmic durations and relationships are different, there is clearly a perceptual relationship between these two examples: they both feature a chromatic ascent, then a descent by leap, followed by descent by step. Following the descent by leap, note duration is increased, and notes occur on points of agogic significance. Since there is a perceptual relationship between these examples, and there is convincing evidence behind Cope's conclusion that many stylistic idioms of composers occur by way of these altered signature shapes, it is worth asking: would the data reduction techniques in this thesis find any relationship between these patterns?

The short answer to this question is no, usually. While some of the recurring ascending chromaticism would show up in *n*-tuple tables for both examples, and the descent by step following a descending leap would show up as entries in the interval transition matrices, there would be no clear logging of the relationship between these two passages. The identification of these passages as significantly similar requires a pattern-matching algorithm which looks for general features, rather than congruent pitch shapes on the score. While the detection of signatures is clearly an ingenious line of investigation into musical style, this paper argues against Cope's declaration that "matching patterns exactly yields little of consequence because precisely repeating sequences are the exception rather than the norm". If the patterns being analysed are of short length, and care is taken so that musical shapes, and not just chromatically identical intervallic patterns, are identified, results are produced which have stylistic significance at a lower level of musical construction than signatures.

There are also examples where signatures *would* be picked up and logged as statistically significant *n*-tuples. Figure 7 shows some more melodic signatures demonstratively used by David Cope, this time from the mazurkas of Chopin. They originate from, in the order as listed here, Op. 24 No. 3, Op. 30 No. 2, Op. 50 No. 3, Op. 33 No. 2, Op. 33 No. 4, and Op. 41 No. 3.

Fig. 7

While it could be argued that the rhythm of these signatures is the most significant part of their shared identity, their consistent shapes are clearly a strong unifying figure. Examples *a, b,* and *c* would be included in 6-tuple tables as (1, -1, -1, -1, -1) and *d, e,* and *f* would occur as (1, -1, -1, -1, 1). Cope lists many variations of this pattern which usually feature an altered 'tail' — sometimes the last three notes descend by thirds, sometimes by third then second, sometimes the six note figure is curtailed to five notes. However, in all these cases, the perceptually strong triplet figure is the same, and in every instance of this pattern, the triplet shape would be included in the 3-tuple tables as (1, -1).

In general, this research makes no effort to link figures of perceptual similarity which are separated by numerous shape and length alterations. Instead, it only looks at the statistical occurrence of very short perceptual shapes, and extrapolates what the occurrence or non-occurrence of these building blocks imply in quantifying musical style.

2.3 - Applications of the Data

Once a piece has been reduced to a single melodic line, and this line is converted to a numerical sequence, statistical operations can be carried out. This thesis proposes the following four means of analysis and comparison which can be applied to the data.

2.3.1 - Most Common Motivic Material

In order to examine recurring motivic material, the melodies (in numerical form) are entered into a program which produces a list of all the unique motives in the data, along with a tally of their incidences. As no limiting heuristics are applied, this program looks at all possible sequential note groupings in a set of melodies. Separate lists are produced for motives of different lengths. In this study, motives of up to six notes are analysed, so five lists are produced of *n*-tuples: a table of 2-tuples, 3-tuples, 4-tuples, 5-tuples and 6-tuples.[5] These lists form almost the entirety of the data used in these four analytical procedures.

The program being used can operate on as much or as little musical material as is required, but statistical significance of results will obviously increase with a larger data input. It is intended that this system of analysis be applied to a corpus of works in order to permit general statements to be made about a composer's stylistic tendencies. However, it can also be used on single works (or even individual sections of works) to compare motivic use in different pieces, rather than just different composers. By comparing the analyses of individual works within a corpus with the analysis of a corpus as a whole, it can be seen which works most strongly reinforce the idiomatic tendencies of a composer.

Even without further analysis, these lists of motives provide insight into compositional idiolects. From them, we can see the most common patterns which occur in a set of works, and compare these between composers. These lists provide empirical reinforcement of traditional musical inferences — for instance, whether or not a particular pitch contour is strongly idiomatic of a composer's output. The most frequently occurring motives in these lists will show the most common building blocks which a composer uses in the creation of phrases, themes, and general melodic figuration.

The significance of these results is best demonstrated by example — figure 8 shows a short passage which is converted into the following numeric sequence: 29.5, 30.5, 31.5, 30.5, 29.5, 33.5, 32.5, 33.5, 34.5, 33.5, 32.5, 31.5, 30.5, 31.5, 32.5, 30.5, 29.5, 30.5, 31.5, 29.5, 33.5, 32.5, 31.5, 30.5, 31.5, 29.5.

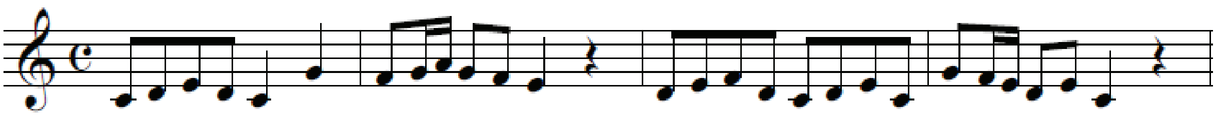


Fig. 8

Here are the 2-tuple, 3-tuple, and 4-tuple motive banks generated from this example:

5 Note that the value n in a given n-tuple is a measure of notes, not intervals. Hence, each n-tuple contains (n-1) intervals — a 2-tuple is the smallest grouping possible, containing one interval between two notes.

2-tuples		
Motive	**Incidences**	**Percentage of total incidences**
-1	11	44.00%
1	9	36.00%
-2	3	12.00%
4	2	8.00%

3-tuples		
Motive	**Incidences**	**Percentage of total incidences**
-1, -1	6	25.00%
+1, +1	4	16.66%
-1, +1	4	16.66%
+1, -2	3	12.50%
+1, -1	2	8.33%
+4, -1	2	8.33%
-1, +4	1	4.16%
-2, -1	1	4.16%
-2, +4	1	4.16%

4-tuples		
Motive	**Incidences**	**Percentage of total incidences**
-1, +1, +1	3	13.04%
-1, -1, -1	3	13.04%
+1, +1, -1	2	8.70%
+1, -1, -1	2	8.70%
-1, -1, +1	2	8.70%
+1, +1, -2	2	8.70%
-1, -1, +4	1	4.35%
-1, +4, -1	1	4.35%
+4, -1, +1	1	4.35%
+1, -2, -1	1	4.35%
-2, -1, +1	1	4.35%
+1, -2, +4	1	4.35%
-2, +4, -1	1	4.35%
+4, -1, -1	1	4.35%
-1, +1, -2	1	4.35%

These figures reflect what an experienced listener might intuitively elicit from the melody in figure 8. The highest incidence figures are either scalic or stepwise enclosures. Towards the bottom of the list (less frequency) are unique figures which mostly occur in non-intuitive division points, such as the 3-tuple (-1, +4) which begins in the middle of the second beat in bar 1. If a motive occurs with sufficient frequency to appear towards the top of the table, it may be relevant despite not appearing at intuitive metrical division points. For instance, in the above example, every ascending fifth is followed by a descending second (+4, -1). This pattern may be viewed as a compositional idiom, despite the fact that this motion occurs over bar-lines and is less obvious than other motives in the passage.

The incidences of unique motives can also be analysed by testing how appropriately they fit various statistical distributions. One such distribution that has found use in musical aesthetics studies in the past is the Zipf distribution (Manaris, Vaughan, Wagner, Romero, & Davis, 2003), which is a type of power law. Zipf's law states that for a set of objects which are ranked by frequency, the occurrence of the object rank i will occur 1/i times the frequency of the most frequent object. This sort of distribution occurs often in physical and social sciences, describing features such as word usage in written texts, or populations in cities. Applied to this study, a Zipfian distribution could mean that in a list of motives, if the most frequent motive appeared 300 times, the second most frequent would appear 150 times, the third most frequent would appear 100 times, and so on.[6] Whether or not a distribution such as this says anything significant about musical aesthetics is a matter of personal opinion. Regardless, if motive occurrence in a given genre of music follows such a distribution, it provides an interesting light on unconscious pattern formation in the human compositional process.[7]

2.3.2 - Interval Transition Matrices

While this thesis primarily examines interesting features in the use of short motives, the program being used can also analyse intervallic transitions. By analysing all of the 3-tuples in a corpus, a table can be made which shows the probability of any transition from one interval to another. This is easily explained through demonstration — the interval transition matrix of

6 Many of the motive-incidence tables shown here feature equal positions; for instance, there are two motives ranked equal first and four motives ranked equal third in the table of 4-tuples in figure 6. Equal positions such as these invalidate a Zipf distribution, but in larger corpus sizes, equal positions occur much lower in the frequency tables and are less relevant.

7 There are hundreds of frequency distributions against which corpus data could be analysed. Zipf's law is isolated here as an example for two main reasons: it finds use in other human artistic corpora (such as literature and MIDI data), and it has been applied to musical analysis in the past (although not in same manner as in this thesis).

the passage in figure 8 is shown below:

From \ To	**-2**	**-1**	**1**	**4**
-2		1(50.00%)		1(50.00%)
-1		6(54.55%)	4(36.36%)	1(9.01%)
1	3(33.33%)	2(22.22%)	4(44.44%)	
4		2(100.00%)		

This table shows the incidences and probabilities for any interval in the left-most column to be followed by the interval in the top-most row. For instance, there are only three descending thirds (-2) that occur in figure 8. One of these is not followed by anything because it is the final interval, so it is not entered into this table. Of the two that are followed by another interval, one descends by a second (-1) and one ascends by a fifth (+4). There is therefore a fifty-percent chance of finding either of those transitions in the analysed material.

It is a valid question to ask whether this table could be used with a random number generator to create pitch patterns which are idiomatic of an analysed corpus of works. This possibility is explored briefly in section 3.2.2. However, it is unlikely that a table such as this could offer any significant stylistic composition tool — without factoring in context, it is unlikely that the results from such a process would produce typical shapes beyond a very short length. Meyer (1957) argued that for statistical data regarding musical events to have meaning and be able to make justifiable predictions, it needs to be employed with a sensitivity to context and musical structures. I agree completely, and the main proposed application of these tables is not to make predictions or produce compositions, but to offer another sort of compositional 'fingerprint' by which composers and works can be compared.[8]

2.3.3 - Measure of Disjunct Writing

By analysing the bank of 2-tuples from a work or corpus, it is possible to make a simple measure of the extent of disjunct writing employed. The approach that has been taken in this study to quantify the disjunctness of a melody is given by the following expression:

$$\frac{\sum_{i \in S} |i| - 1}{n}$$

8 It is worth mentioning that the use of a transition matrix in analysing music is not at all new. Youngblood(1958) proposed a matrix showing probabilities of transitions, but used scale degrees (for instance, probability of a supertonic moving to submediant) rather than isolated intervallic movement. Almost every work testing the 'implication-realization' model uses interval transition tables of some kind.

where S is a set containing all intervals in a corpus, *i* are all the elements of *S* which are greater than +1 or less than -1, and *n* is the total amount of elements in *S*. In other words, for each interval in the analysed corpus which has an absolute value greater than 1 (which means it appears on the staff as a third or wider), the portion of the absolute value greater than 1 is summed, and this total sum of these values is then divided by the total number of intervals in the corpus. In effect, this averages the disjunct value of every note. In this thesis, this value is called the 'disjunct writing score'.

While the disjunct writing score has no explicit meaning in itself (besides that outlined above), it is a means of comparison between composers. Since it is averaged over all notes in a corpus, the size of the set S does not skew the results as it scales. Both larger intervals and more frequent disjunct motion increase this score. The minimum disjunct writing score is 0, for a corpus which features no movement greater than a step. The maximum disjunct writing score is equal to the widest interval possible in the medium being analysed, minus 1. These values mean very little in isolation, and their application is best shown with a comparison. The disjunct writing score of the passage in figure 8 is 0.36. This excerpt is almost entirely scalic, and hence gives a very low score. Compare this to the disjunct writing score of the first four bars of the Mozart *Serenade* in figure 1 (which contains only one conjunct interval), which is 1.41. This score is an efficient one-figure summary of how disjunct the melodic writing is within a corpus.

It is worth noting that a graph of intervals in a corpus and their frequency reveals a more thorough description of the nature of disjunct writing employed. This will be discussed with the results in section 3.2.3.

2.3.4 - Motivic Concentration

In this thesis, 'motivic concentration' is a term used to describe the extent to which motives are re-used. A work which has a small number of motives which are re-used often has a high level of motivic concentration. Conversely, a work which uses many different motives, and has fewer which are re-used heavily, has a low level of motivic concentration.[9] This is something which is less easy to quantify objectively. However, there are still some easily quantifiable ratios within the results tables which, while not foolproof, will usually generate a

9 The other two theoretical constructs (many motives used heavily, few motives used sparsely) are feasible, but practically non-existant within musical corpora. All real examples fall somewhere between these four cardinal extremes.

useful metric for comparison. Consider the table of 3-tuples derived from figure 8, and the table of 3-tuples from figure 9, a contrasting passage:

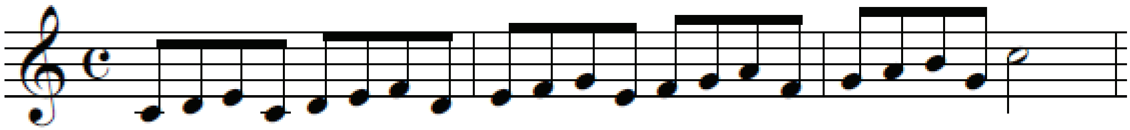


Fig. 9

3-tuples from Figure 9		
Motive:	**Incidences:**	**Percentage of total incidences:**
+1, +1	9	47.37%
+1, -2	5	26.31%
-2, +1	4	21.05%
-2, +3	1	5.26%

While figure 9 has slightly fewer notes, it has dramatically fewer unique motives than figure 8. The reason why is obvious — figure 9 is simply a repeating four-note figure, while figure 8 features more variation. The table of 3-tuples for figure 9 shows that almost 95% of the total material in the excerpt can be created using only three motives. While Figure 8 also has a relatively high level of motivic concentration, eight motives are required to create 95% of the material. We can also see that in figure 8, there are nine unique motives, with twenty-four total 3-tuples in the excerpt. This is a higher ratio than figure 9, which has four unique motives, with nineteen total 3-tuples in the excerpt. These general observations suggest that by considering a corpus's total unique *n*-tuples, total *n*-tuples, and proportion of motives needed to account for a particular percentage of the corpus, it is possible to create some sort of numerical quantifier of motivic concentration. While discussion of results with the use of full tables and graphs is obviously going to be more enlightening as to the full nature of motivic concentration in a corpus, the following two metrics for comparison are proposed (these are all in reference to the 3-tuple tables only):

- Unique Motives over Length (UML score): This score is simply the total number of unique motives divided by the total amount of motives in a corpus. In the above examples, the UML score of figure 8 is 0.374, and the UML score of figure 9 is 0.21. A lower score implies higher motivic concentration, since it is lowered by having fewer total unique motives, and raised by having more unique motives. The highest theoretical score is 1, for a corpus which contains no repeated motives. The UML score approaches a limit of 0 as the total number of motives increases (assuming a static number of unique motives).

- Motives to Cover *n*% (MC*n*): This value is the minimum amount of unique motives required to account for *n*% of the material. The value of *n* is variable because corpora of different sizes may require different values of *n* to make meaningful comparisons. In the above example, figure 8 has a MC50 value of 3 (the first three motives in the table account for 58.33% of the material) while figure 9 has an MC50 value of 2 (the first two motives in the table account for 73.68% of the material). It is also illuminating to see that there is a bigger contrast between excerpts when the value of *n* is higher: MC90 in figure 8 is 7, while in figure 9 it is 3. This generally suggests that for any value of *n* used to compare two corpora, the corpus with the smaller MC*n* score is probably more motivically concentrated. When comparing corpora larger than single excerpts, smaller values of *n* are required to see meaningful results. Because the amount of unique motives used will naturally increase as the amount of entries in the corpus grows, comparisons between corpora of very uneven sizes will produce confusing results — smaller corpora will tend to appear to be more motivically concentrated. Despite this, the MC*n* value is a meaningful form of comparison between corpora which are of similar size.

There are other factors which can be assessed subjectively in discussing motivic concentration — graphs showing motives against incidences produce curves which demonstrate motivic concentration much more thoroughly than a single number. Such graphs are shown in section 3.2.4. It is also important to remember that a different set of results will be produced for each *n*-tuple table — the values above only examine three-note figures. Section 3.2.4 discusses results and values for 3-tuples, 4-tuples, 5-tuples and 6-tuples. 2-tuples are not addressed in terms of motivic concentration, because they represent only single intervals, and do not merit the same sort of analysis as longer musical elements. The re-use in a work or corpus of a small number of intervals, or a wide variety of intervals, is instantly recognisable upon examination of the 2-tuple tables. The applicability of different length *n*-tuples depends on attributes of the corpora being analysed. One example which shows how this is the case is the consideration of metre: due to the frequency of typical shapes and rhythmic groupings, compound time signatures will often feature more repeated three and six note patterns than simple time signatures.

Motivic concentration is much more difficult to identify and label with a single figure than the other attributes which have been outlined. Despite this, by comparing motive frequency graphs, along with UML and MC*n* values, a meaningful comparison between composers' stylistic reuse of pitch material is made apparent.

3 - Application to Classical Repertoire

Classical-era melodies feature frequent reuse and development of small pitch-based motives, and thus form an ideal experimental environment for this research. The experimental application of the systems outlined in previous sections of this thesis is here applied to twelve piano sonata expositions. From this application, compositional features are compared, and conclusions are drawn regarding the differences in style between individual works and composers.

3.1 - Choice of Repertoire

This research looks at the expositions from the first movements of the following piano sonatas:

By Wolfgang Amadeus Mozart:

- Sonata No. 1 in C Major (1774)
- Sonata No. 6 in D Major (1775)
- Sonata No. 7 in C Major (1777)
- Sonata No. 13 in B♭ Major (1783)

By Muzio Clementi:

- Sonata Op. 1 No. 2 in G Major (1771)
- Sonata Op. 1 No. 4 in F Major (1771)
- Sonata Op. 2 No. 4 in A Major (1781)
- Sonata Op. 8 No. 3 in B♭ Major (1782)

By Joseph Haydn:

- Sonata No. 37 in E Major (1773)
- Sonata No. 46 in E Major (1776)
- Sonata No. 50 in D Major (1780)
- Sonata No. 51 in E♭ Major (1780)

These works have been chosen because they fit a common set of criteria. They are all in major keys, all in common time, and feature roughly similar tempo markings (moderato and faster). Historically, they all fit within a compositional period of about 20 years (around the 1770s to 1780s). Also, they all feature melodic lines which are mostly confined to the highest sounding voice, and their texture is almost entirely homophonic. By controlling these variables, and selecting works of similar characteristics, it is intended that any comparisons drawn are between the composers' idiolects, and not due to other factors such as differing compositional forms or differing historical trends.

It is acknowledged that each of these composers' compositional styles changed over the course of their life. The data from this analysis allows only limited comparison between composers — the corpora represent only a small selection of writing from immense collections of works. As such, the conclusions formed, while potentially allowing more general extrapolation, can only directly apply to a very small portion of each composer's output. The results discussed in section 3.2 therefore do not aim to compare the melodic idiolects of these three composers in a general sense, but draw conclusions about the nature of their piano melody writing during the late eighteenth century.

3.1.1 - Melody or Figuration?

This study uses works which predominantly feature melodic lines occurring in the highest sounding voice in their texture. Because of this, with very few exceptions, the top sounding voice at any point in the music can be taken as the principal melodic line.[10] This approach does run into some complications due to inner-voice melodic figures and typical classical-era embellishing figuration. While every case needs to be dealt with individually, the general approach followed in this thesis is discussed with the following examples: figure 10 is the third bar from Haydn's Sonata No. 40 in E♭ (which isn't used in this experiment, and is just shown here for demonstration purposes), and figure 11 is a passage from Clementi's Sonata No. 3 from the Op. 8 set.

10 In most cases. Obviously, if a bass or accompaniment figure continues through a short rest in the melody, the melodic line is still the top voice.

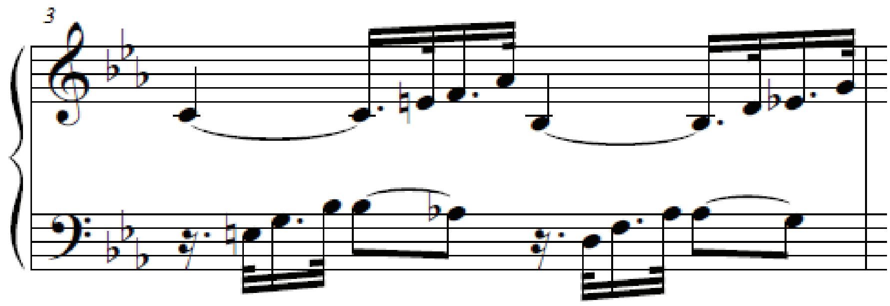


fig. 10

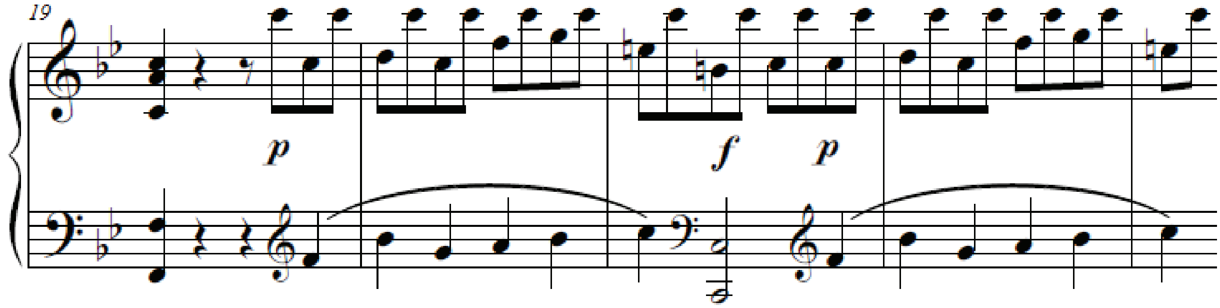


fig. 11

In most cases in the piano repertoire of these three composers, when there are two upper voices which are moving melodically, it is clear that the bottom voice is functioning in a purely harmonic capacity. The harmonisation of melodic figures in thirds or sixths is extremely common. In these cases, the perceptual significance of the top voice is reason enough to exclude the bottom voice from a melodic reduction.[11] Figure 10 represents a potential complication for the reduction process previously outlined — in a work which is predominantly homophonic, this bar clearly implies a duplicity of melodic voices. The dotted rhythm arpeggio figure is shared between the two voices in a similar style to that of a two part invention. It is not appropriate to consider that the bar contains a single melody which jumps between registers on each beat because the top voice sustains during beats one and three and the bottom voice is still active in beats two and four — both voices are continuous. For passages of limited polyphony such as this which occur in the works being analysed, the top voice alone is reduced, and the bottom voice is ignored. While this is partially justified by the perceptual significance of top voices in music, this is still an unfortunate concession which must be made in the process of monodic reduction. However, it should be noted that the works chosen for analysis in this thesis feature very few significant polyphonic passages.

11 It is treated as axiomatic that the outer (and most active) voices in a texture are the most easily perceived. This is supported by the notion of perceptual streaming in auditory psychological research — test subjects tend to organise pitch fields of general continuous register into disparate lines, as discussed by Deutsch (1991). It stands to reason that an upper melodic line, which has the largest frequency spectrum range to itself would suffer the least from auditory 'masking' and therefore be the most easily perceptible. However, it should be noted that algorithmic research which models melody extraction on human perception most strongly correlates pitch salience with melody (Paiva, 2005). This suggests that any line could be made perceptually melodic by the decisions of a skilled performer.

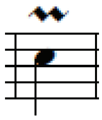
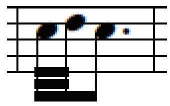
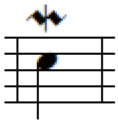
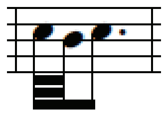
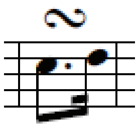
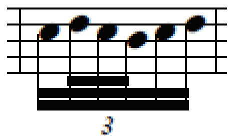
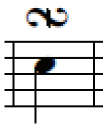
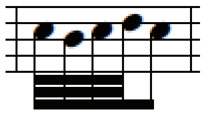
Another problematic issue which forces us to reconsider our definition of melody is demonstrated in figure 11. Here, the top voice is embellishing the harmonic movement in the left hand with a quaver pattern which suggests an inverted pedal point. This line does not possess many of the characteristics which musicologists would intuitively associate with a melody — its rhythm is constant, and if the voicing between the treble and bass were reversed, it would clearly form an accompanying line. However, there is still obviously some significance to this being the top voice in the texture, and the question of where the melody ends and figuration begins is extremely subjective. In these analyses, figuration such as this is treated as melodic, and taken in the melodic reduction with all the other top voice material. This significantly reduces the subjectivity involved in producing a monodic reduction, but requires a broadening of what is encapsulated under the term 'melody'. If figuration such as the above is included, then it must be stated that this thesis is not only analysing melodic idioms in piano writing, but idioms of all top voice figuration in the piano corpora being analysed, melodic or otherwise.

A full listing of the passages in this study which were problematic to reduce to monophony are listed in the appendix, along with details of how they are dealt with in the reductions.

3.1.2 - On Ornamentation

Ornamentation, both explicitly written or implied with symbols, is an important part of melody writing in many musical genres. Since this study intends to look at melodic idioms, and ornamentation is a part of melodic writing, ornamentation should be included in the reduction. An obvious problem emerges with the open-ended nature of ornament interpretation. As any specialist in early music can attest, how an ornament is performed depends on compositional context, historical and geographical context, and a performer's own tastes and musical intuitions. In converting an ornament on the score to a series of pitches for the purposes of this analysis, a subjective decision is being made. This is acknowledged as a concession to subjective interpretation for the purpose of representing ornamental figuration in the analysis.

While all ornaments could be left out of the analysis, this would exclude many figures which may have motivic significance. If turns, mordents, or similar figures occur frequently melodically, then some effort should be made to represent these figures in motivic frequency tables. While the interpretation of turns and trills may imply numerous possible sequences of pitches, some ornaments, such as mordents, acciaccaturas, and appoggiaturas tend to have a relatively standardised pitch interpretation, even if they are rhythmically open to interpretation (Blood, 2011). No attempt is made to generalise the entry of all ornamental material for this system of reduction, but for the repertoire examples used in this experiment, the following conventions are used to maintain consistency (note that the rhythmic interpretations are shown purely for example and are irrelevant to the reduction):

Ornament symbol		**Data format interpretation**
	Mordents are uniformly interpreted as the initial note, the note in the present key immediately above the initial note, and a return to the initial note.	
	Inverted mordents are uniformly interpreted as the initial note, the note in the present key immediately below the initial note, and a return to the initial note.	
	Turns, whether they are placed on a note or between two notes, are interpreted as a diatonic enclosure pattern consisting of five notes.	
	This figure, sometimes called a half-mordent, occurs in the works of Haydn and is treated as a openly interpretable curiosity. In parts of the Vienna Urtext edition of his piano works where it has been editorially changed to a mordent or other conventional sign, the editorial mark is used. In situations where it is left as is, it is interpreted here as an inverted turn (primarily to differentiate it from the traditional turn figure). Note that this interpretation is still relatively arbitrary, but of relatively minor significance due to the low occurrence of this figure.	

All trills which are indicated by symbols are ignored. This is because they are mostly not used motivically, instead finding use as typical cadential embellishment or highlighting structural divisions. Attempting to standardise any interpretation of trills is a far more subjective process than with other ornaments, and would introduce many extraneous entries into the motive tables which would potentially skew the results. Despite this, any figuration which could be considered embellishing but is written out in pitches is included in the analysis like any other material. While also termed as 'ornamentation', appoggiaturas and acciaccaturas present no problems in melodic reduction as they are closed in terms of pitch.

3.2 - Presentation and Analysis of Results

3.2.1 - Most Common Motivic Material

The tables of incidences of motives for each set of four works are far too long to list in their entirety here — the 6-tuple incidence table for the Mozart corpus, for instance, is 1008 lines long. For the sake of concision, the incidence-tables here feature only the thirty highest ranked *n*-tuples.[12] 2-tuple tables will be discussed in section 3.2.3, because they summarise musical motion in general, rather than specific motives.

The thirty highest ranked *n*-tuples (with incidences and percentage of total material accounted for by each motive) from the 3-tuple, 4-tuple, 5-tuple, and 6-tuple banks are shown on the following pages:

12 All n-tuple lists for all corpora and all individual expositions are included on the accompanying compact disc. It should also be noted that these tables have been truncated to the first thirty entries, ignoring equal rankings. Equally ranked motives are placed in the lists in the order that they occur in the analysed corpora, so the inclusion and exclusion of some equally ranked motives is arbitrary.

3-tuple Incidences								
Clementi			**Haydn**			**Mozart**		
Motive	**Incid.**	**Perc.**	**Motive**	**Incid.**	**Perc.**	**Motive**	**Incid.**	**Perc.**
-1, -1	178	13.44	-1, -1	263	18.37	-1, -1	353	19.07
1, 1	171	12.92	1, 1	133	9.29	1, 1	161	8.7
-1, 1	134	10.12	-1, 1	114	7.96	2, -1	108	5.83
1, -1	112	8.46	1, -1	91	6.35	-1, 1	99	5.35
0, -1	27	2.04	0, -1	46	3.21	1, -1	89	4.81
1, -3	25	1.89	-1, 0	46	3.21	-1, 2	75	4.05
-2, 1	24	1.81	2, -1	31	2.16	1, -2	49	2.65
1, -2	21	1.59	0, 0	25	1.75	-2, 1	46	2.49
2, -1	20	1.51	-1, 2	24	1.68	-1, 0	44	2.38
-1, 0	20	1.51	-2, -2	23	1.61	-2, -2	44	2.38
0, 1	18	1.36	-2, -1	23	1.61	0, -1	42	2.27
3, -3	16	1.21	1, -2	22	1.54	1, 2	31	1.67
2, -2	16	1.21	2, 2	20	1.4	-1, -2	26	1.4
1, 2	15	1.13	-2, 1	18	1.26	0, 0	22	1.19
-2, -2	15	1.13	0, 1	17	1.19	2, 2	20	1.08
-1, 2	14	1.06	3, -3	14	0.98	-2, -1	18	0.97
-7, 7	14	1.06	1, 2	13	0.91	2, 3	17	0.92
-3, 1	13	0.98	2, 3	13	0.91	5, -1	17	0.92
-3, -1	12	0.91	-1, -2	12	0.84	1, 0	16	0.86
-3, -2	12	0.91	3, 2	12	0.84	0, 1	16	0.86
0, 0	12	0.91	2, -2	12	0.84	-1, 5	15	0.81
-3, 3	11	0.83	3, -1	11	0.77	-3, -2	15	0.81
5, -5	11	0.83	-3, 2	11	0.77	2, 1	13	0.7
-3, 0	10	0.76	-1, 3	11	0.77	3, 2	13	0.7
-2, 0	10	0.76	-3, -1	10	0.7	-2, 2	13	0.7
-2, -3	10	0.76	-3, -2	10	0.7	-2, -3	13	0.7
-2, -1	9	0.68	-2, 3	9	0.63	2, -2	12	0.65
1, 3	9	0.68	4, -4	9	0.63	0, -2	12	0.65
6, -6	9	0.68	5, -2	9	0.63	1, -3	11	0.59
-1, 7	8	0.6	1, 0	8	0.56	-5, 5	11	0.59

4-tuple Incidences								
Clementi			**Haydn**			**Mozart**		
Motive	**Incid.**	**Perc.**	**Motive**	**Incid.**	**Perc.**	**Motive**	**Incid.**	**Perc.**
1, 1, 1	100	7.58	-1, -1, -1	162	11.34	-1, -1, -1	228	12.34
-1, -1, -1	82	6.21	-1, -1, 1	57	3.99	1, 1, 1	81	4.39
-1, -1, 1	67	5.08	1, 1, 1	55	3.85	-1, 2, -1	61	3.3
-1, 1, -1	64	4.85	-1, 1, 1	51	3.57	-1, -1, 1	49	2.65
1, -1, -1	58	4.39	1, 1, -1	48	3.36	1, -1, -1	46	2.49
1, 1, -1	35	2.65	1, -1, -1	43	3.01	2, -1, 2	43	2.33
-1, 1, 1	29	2.2	-1, 1, -1	31	2.17	-1, 1, 1	38	2.06
1, -1, 1	29	2.2	-1, 0, -1	27	1.89	2, -1, -1	37	2
1, 1, -3	11	0.83	1, -1, 1	22	1.54	1, 1, -1	36	1.95
0, -1, 1	10	0.76	0, -1, 0	19	1.33	-1, -1, 0	31	1.68
-1, -1, 0	10	0.76	-1, 2, -1	17	1.19	-1, 1, -1	26	1.41
1, -2, 1	10	0.76	0, -1, -1	14	0.98	1, -1, 1	19	1.03
-3, 1, 1	10	0.76	-2, -1, 1	14	0.98	1, 2, -1	17	0.92
-1, 1, -3	9	0.68	2, -1, -1	13	0.91	-1, 0, -1	17	0.92
-1, 2, -1	9	0.68	0, 0, 0	13	0.91	0, -1, -1	17	0.92
2, -1, -1	9	0.68	1, -2, -1	10	0.7	-2, -2, 1	17	0.92
0, -1, -1	9	0.68	1, 1, -2	10	0.7	1, 1, -2	16	0.87
-1, 1, 2	9	0.68	1, -1, 0	9	0.63	-2, 1, 1	14	0.76
0, 1, 1	9	0.68	3, -1, -1	9	0.63	-1, 1, -2	14	0.76
-2, 1, 1	8	0.61	-1, 1, 3	8	0.56	1, -2, -2	14	0.76
-1, 0, 1	8	0.61	1, -2, 1	8	0.56	-1, 5, -1	12	0.65
2, 0, -1	7	0.53	2, -1, 2	8	0.56	1, -2, 1	12	0.65
1, 1, -2	7	0.53	-1, 1, 2	7	0.49	2, -1, -2	12	0.65
1, -3, 1	7	0.53	-2, 1, 1	7	0.49	-1, 1, 2	11	0.6
-2, 1, -2	7	0.53	-1, -1, 2	7	0.49	-1, -1, 2	11	0.6
1, -3, 0	6	0.45	-3, -2, -2	7	0.49	-1, -1, 5	11	0.6
-2, 0, -1	6	0.45	-7, 0, 0	7	0.49	-3, -2, -2	11	0.6
-1, -1, 2	6	0.45	-1, -1, 4	6	0.42	-1, -2, 1	10	0.54
-1, 0, -1	6	0.45	-1, -1, 0	6	0.42	2, 3, 2	10	0.54
1, -1, 7	6	0.45	3, 2, 2	6	0.42	2, 2, 3	10	0.54

5-tuple Incidences								
Clementi			**Haydn**			**Mozart**		
Motive	**Incid.**	**Perc.**	**Motive**	**Incid.**	**Perc.**	**Motive**	**Incid.**	**Perc.**
1, 1, 1, 1	57	4.33	-1, -1, -1, -1	101	7.09	-1, -1, -1, -1	134	7.27
-1, -1, -1, -1	52	3.95	-1, -1, 1, 1	32	2.25	1, 1, 1, 1	41	2.22
-1, -1, 1, -1	41	3.12	-1, -1, -1, 1	31	2.18	2, -1, 2, -1	36	1.95
1, -1, -1, 1	39	2.96	1, 1, 1, 1	31	2.18	-1, 2, -1, 2	30	1.63
-1, 1, -1, -1	32	2.43	-1, 1, 1, -1	29	2.04	2, -1, -1, -1	29	1.57
1, 1, -1, -1	21	1.6	1, -1, -1, -1	29	2.04	1, -1, -1, -1	29	1.57
-1, 1, -1, 1	20	1.52	1, 1, -1, -1	24	1.69	-1, -1, -1, 1	28	1.52
-1, -1, 1, 1	15	1.14	-1, 1, -1, 1	17	1.19	-1, -1, -1, 0	26	1.41
-1, -1, -1, 1	15	1.14	-1, 0, -1, 0	13	0.91	1, 1, -1, -1	24	1.3
1, 1, 1, -1	14	1.06	-1, -1, 1, -1	13	0.91	-1, -1, 1, 1	19	1.03
-1, 1, 1, 1	13	0.99	-1, 1, 1, 1	12	0.84	-1, 1, 1, 1	19	1.03
-1, 1, 1, -1	13	0.99	1, -1, -1, 1	11	0.77	-1, 2, -1, -1	18	0.98
1, -1, -1, -1	12	0.91	0, -1, 0, -1	11	0.77	-1, -1, 1, -1	16	0.87
1, 1, 1, -3	11	0.84	1, 1, 1, -1	11	0.77	1, 1, 1, -1	15	0.81
1, -1, 1, 1	9	0.68	-1, 1, -1, -1	10	0.7	0, -1, -1, -1	14	0.76
0, -1, 1, -1	8	0.61	1, 1, -1, 0	9	0.63	-1, 1, 1, -1	14	0.76
-3, 1, 1, 1	8	0.61	-1, 0, -1, -1	8	0.56	-1, -1, 0, -1	13	0.71
1, 1, -1, 1	7	0.53	1, -1, 1, 1	8	0.56	1, -1, -1, 1	12	0.65
1, -1, 1, -1	7	0.53	-1, 2, -1, -1	8	0.56	-1, 1, -1, -1	11	0.6
1, 1, -3, 1	7	0.53	-1, 1, 1, -2	8	0.56	1, 2, -1, -1	10	0.54
1, -3, 1, 1	7	0.53	2, -1, 2, -1	8	0.56	-1, -1, -1, 5	10	0.54
-1, -1, 1, -3	6	0.46	0, 0, 0, 0	8	0.56	1, -2, -2, 1	10	0.54
-1, 1, -3, 0	6	0.46	0, -1, -1, -1	7	0.49	-1, -1, 1, 2	9	0.49
-1, 2, -1, -1	6	0.46	1, -1, 1, -1	7	0.49	5, -1, -1, -1	9	0.49
-2, 1, -2, 1	6	0.46	1, -2, -1, 1	7	0.49	-2, 1, 1, 1	9	0.49
1, -1, 0, -1	6	0.46	0, -1, -1, 1	6	0.42	-1, -1, -1, 2	8	0.43
-1, 0, 1, 1	6	0.46	1, -1, 0, -1	6	0.42	-1, -1, 2, -1	8	0.43
-1, -1, 2, -1	5	0.38	-1, -1, -1, 4	6	0.42	-1, -1, 5, -1	8	0.43
0, -1, -1, 1	5	0.38	3, -1, -1, -1	6	0.42	-1, 5, -1, -1	8	0.43
-1, -1, -1, 0	5	0.38	-2, -1, 1, 1	6	0.42	1, 1, -2, -2	8	0.43

6-tuple Incidences								
Clementi			**Haydn**			**Mozart**		
Motive	**Incid.**	**Perc.**	**Motive**	**Incid.**	**Perc.**	**Motive**	**Incid.**	**Perc.**
1, 1, 1, 1, 1	39	2.97	-1, -1, -1, -1, -1	73	5.14	-1, -1, -1, -1, -1	76	4.13
-1, -1, -1, -1, -1	35	2.67	1, 1, 1, 1, 1	21	1.48	-1, 2, -1, 2, -1	26	1.41
1, -1, -1, 1, -1	32	2.44	-1, -1, 1, 1, -1	20	1.41	1, 1, 1, 1, 1	23	1.25
-1, -1, 1, -1, -1	25	1.91	1, 1, -1, -1, -1	14	0.99	-1, -1, -1, -1, 1	21	1.14
-1, 1, -1, -1, 1	23	1.75	-1, -1, -1, 1, 1	14	0.99	2, -1, 2, -1, 2	19	1.03
1, 1, -1, -1, 1	13	0.99	-1, -1, -1, -1, 1	13	0.92	2, -1, -1, -1, -1	19	1.03
-1, -1, 1, -1, 1	12	0.91	1, -1, -1, -1, -1	13	0.92	1, -1, -1, -1, -1	19	1.03
-1, 1, 1, -1, -1	10	0.76	-1, 1, 1, -1, -1	12	0.85	1, 1, -1, -1, -1	18	0.98
1, 1, 1, 1, -1	9	0.69	-1, -1, 1, 1, 1	10	0.7	-1, 2, -1, -1, -1	17	0.92
-1, -1, 1, 1, 1	8	0.61	-1, 1, -1, -1, -1	10	0.7	-1, -1, 1, 1, 1	13	0.71
1, -1, -1, -1, -1	7	0.53	1, -1, -1, 1, 1	9	0.63	-1, -1, -1, 0, -1	12	0.65
-1, -1, -1, -1, 1	7	0.53	0, -1, 0, -1, 0	9	0.63	-1, 1, 1, -1, -1	12	0.65
-1, 1, -1, 1, 1	7	0.53	1, 1, 1, -1, -1	9	0.63	-1, -1, -1, -1, 0	12	0.65
1, 1, 1, -3, 1	7	0.53	-1, -1, -1, 1, -1	9	0.63	-1, -1, -1, 1, 1	12	0.65
1, 1, -3, 1, 1	7	0.53	-1, 1, 1, -1, 0	8	0.56	-1, -1, -1, 1, -1	11	0.6
-1, -1, 1, -3, 0	6	0.46	-1, 0, -1, 0, -1	8	0.56	2, -1, 2, -1, -1	10	0.54
1, 1, 1, -1, 1	6	0.46	-1, -1, 1, -1, -1	8	0.56	5, -1, -1, -1, -1	9	0.49
-1, -1, -1, 1, 1	6	0.46	1, -1, -1, -1, 1	7	0.49	1, 1, 1, -1, -1	9	0.49
-1, -1, -1, 1, -1	6	0.46	1, 1, -1, -1, 1	7	0.49	-1, -1, -1, -1, 5	8	0.44
1, 1, 1, -1, -1	6	0.46	-1, 1, 1, 1, 1	7	0.49	-1, 5, -1, -1, -1	8	0.44
1, -1, 0, 1, 1	6	0.46	0, 0, 0, 0, 0	7	0.49	-1, 1, 1, 1, 1	8	0.44
1, -3, 1, 1, 1	6	0.46	1, 1, -1, 0, -1	6	0.42	1, 1, 1, 1, -1	8	0.44
-3, 1, 1, 1, -3	6	0.46	1, -1, 0, -1, -1	6	0.42	0, -1, -1, -1, 0	7	0.38
-1, -1, 1, 1, -1	5	0.38	1, -1, 1, 1, -1	6	0.42	-1, -1, -1, 5, -1	7	0.38
-1, 1, -1, -1, -1	5	0.38	-1, 1, -1, 1, -1	6	0.42	1, 1, -2, -2, 1	7	0.38
1, 1, -1, -1, -1	5	0.38	1, -1, 1, -1, 1	6	0.42	-1, -1, -1, 2, -1	6	0.33
-7, 1, 1, 1, 1	5	0.38	-1, 2, -1, 2, -1	6	0.42	1, 2, -1, -1, -1	6	0.33
1, -2, 1, -2, 1	5	0.38	-1, 1, 1, 1, -1	5	0.35	6, -6, 6, -6, 6	6	0.33
0, -1, 1, -1, 1	5	0.38	-2, -1, 1, 1, -2	5	0.35	-5, 5, -5, 5, -5	6	0.33
-1, 1, -1, 1, 2	5	0.38	1, 1, 1, 1, -1	5	0.35	5, -5, 5, -5, 5	6	0.33

These tables show the extent to which the corpora under analysis reuse short pitch-based motivic fragments. Naturally, since these works are all chosen to feature similar characteristics, the motivic data extracted from these corpora shows many similarities. Many of the highest scoring motives are shared between the corpora. These similarities are largely accounted for by idioms of eighteenth-century music theory — Robert Gjerdingen discusses at length in *Music in the Galant Style* how classical-era composers were educated in the use of a collection of typical recurring melodic figures (Gjerdingen, 2007). As a theorist writing in the 1730s, Johann Mattheson describes melodic formulation as the reuse and manipulation of short, typical motives (Lester, 1992, pp. 158-174). This recurring figuration is shown in many of the highest ranked motives in the tables. Each table shows melodic figures typical of the length being analysed; for instance, all 3-tuple tables feature mordent (1, -1) and inverted mordent (-1, 1) figures within the first five entries, while 6-tuple tables show mostly varied scalic passages in their thirty highest entries. 4-tuple tables show typical four-note ornaments or figuration, such as turns (-1, 1, 1) or incomplete neighbour note enclosures (-1, 2, -1). Other commonly appearing melodic figures which are represented with relatively high rankings in the tables include scales in thirds (-1, 2, -1, 2), appoggiatura sequences (-1, 0, -1 or similar) and arpeggiation figures (2, 2 or -2, -2 or 2, 3 or similar).

A particularly striking feature of all the material in the tables is the predominance of scalic writing, especially descending motion. In both the Mozart and Haydn corpora, descending scales are the highest ranked motives by far, and in the Clementi corpus, ascending and descending scales appear at the top of the lists nearly equally. While the corpora being examined here are too small to make generalisations about larger musical genres, this analysis confirms what is found in the previous study by Vos — that Classical music has the broad tendency to feature contours which ascend by leap and descend by step (1989). Comparison with other fields drawing this conclusion occurs in section 3.3.1.

While there are many obvious similarities between these corpora, these tables also allow the examination of what is different. In a now famous review of Beethoven's Fifth Symphony, composer and critic E. T. A. Hoffman described the output of Haydn as being ‘dominated by childlike optimism’, in contrast to Mozart's, which ‘leads us deep into the realm of spirits’ (Hoffmann, 1989). It is a valid question to ask whether such contrasting emotive perceptual responses are at all justified by varied patterns of motivic use. This thesis makes no attempt to engage this question — instead, curious dissimilarities between corpora will be discussed, and the implications of these results for the fields of perception or psychology are left to future scholars.

The most striking contrast between the three corpora has already been stated — where the Mozart and Haydn works have descending scalic movement as their highest ranked motive in every bank, the Clementi corpus has ascending and descending scales ranked relatively equally. This seems to reflect a difference in the 'ascent by leap, descent by step' relation previously mentioned. Out of the three corpora, the Clementi expositions appear to feature the most occurrences of embellishing figuration in scalic form (as opposed to arpeggio figures).

While many of the motives in these tables are shared, there is a significant curiosity in the works of Mozart which is highlighted by examining the banks of 4-tuples. There are eight possible combinations of ascending and descending seconds in a four-note group.[13] All of these eight combinations appear within the first eight entries of the Clementi corpus, and as entries 1-7, and 9 of the Haydn corpus. Mozart's tables seem to reflect a greater use of figures containing leaps: these eight figures are contained within the first twelve 4-tuples, and the first non-scalic motive is ranked third, much higher than in the Clementi and Haydn corpora. This pattern also occurs in the other banks: in the 6-tuple banks, the first disjunct Clementi motive is ranked 14th, the first disjunct Haydn motive 27th, and the first disjunct Mozart motive much higher up in 2nd position. While they are not shown in the truncated tables presented here, the motives in the Mozart 6-tuple table in ranks 31 and onwards feature many more disjunct intervals than entries of similar rank in the Clementi or Haydn tables. Fuller discussion of conjunct and disjunct motion is reserved for section 3.2.3, but it seems highly significant that Mozart's corpus demonstrates a predilection for disjunct motion in the material most frequently used.

In section 2.3.1, it was mentioned that motivic frequencies could be tested to fit statistical distributions. While none of the motive banks shown so far follow precisely Zipfian distributions, they definitely follow the general shape of a power law distribution. Consider figure 12, a graph which shows the distribution of rank against frequency in the Haydn 3-tuple table:

13 As outlined in information theory — with two possible states (1, -1) and three intervals in a 4-tuple, the total combinations possible are 2^3. More generally, with n possible states and i intervals, there are n^i combinations.

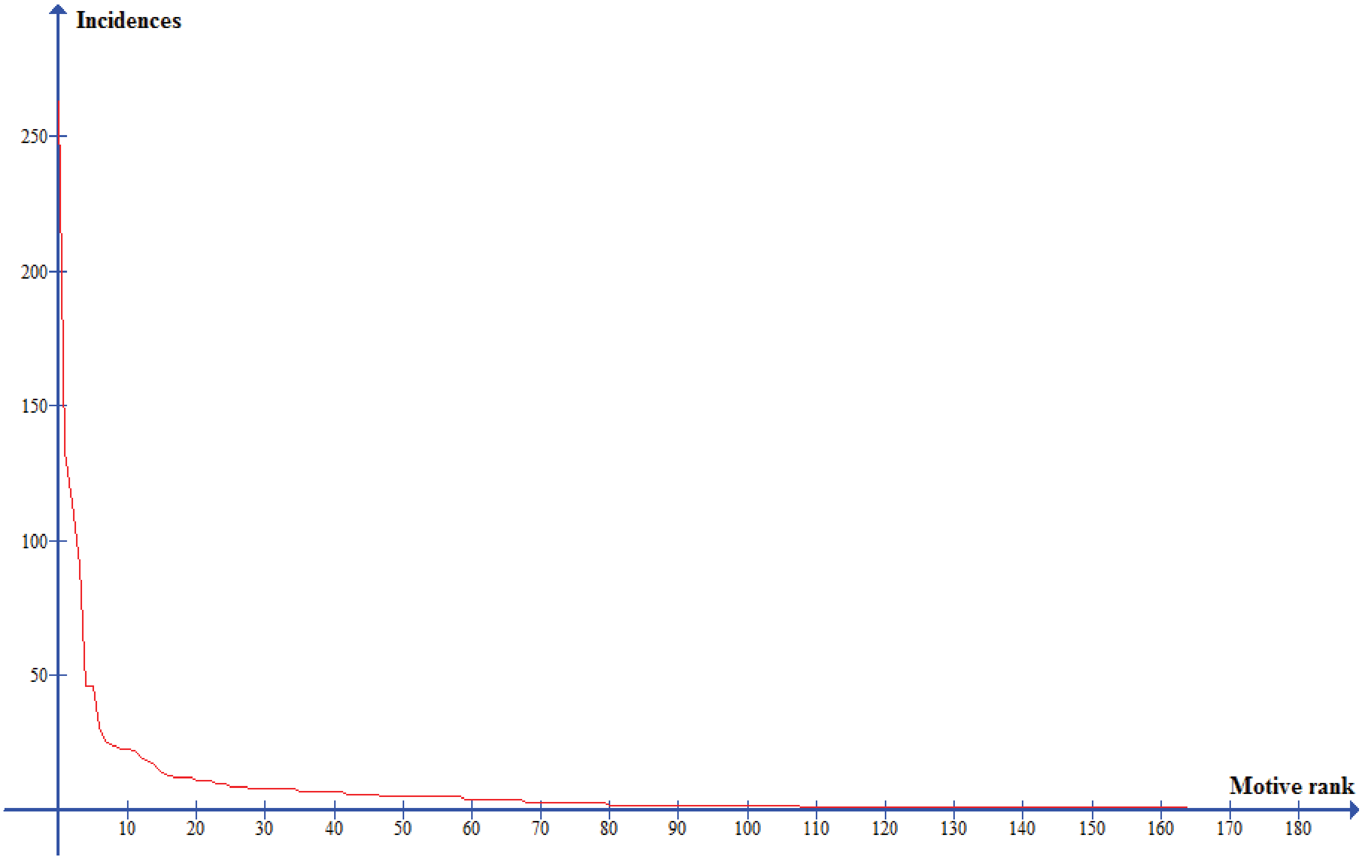


Fig. 12

Due to the fact that all banks feature many motives with only one or two incidences, this distribution is made clearer on a graph where the *x*-axis is displayed logarithmically, as in figure 13:

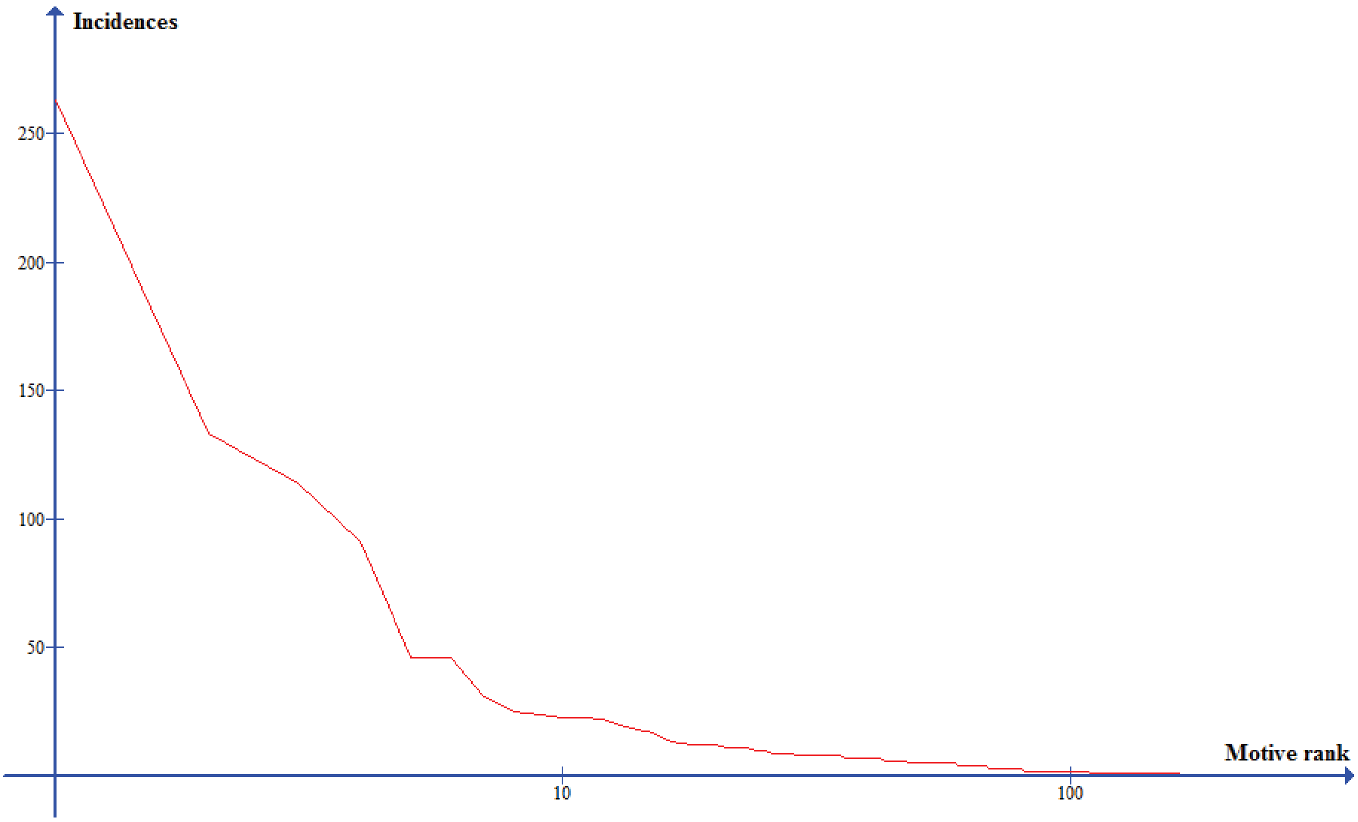


Fig. 13

Section 3.2.4 looks at these incidence distributions in more detail. This graph is simply one typical example which demonstrates the general trend in these results — one or two motives in the highest ranking positions which occur dramatically more than any other, a gradual decrease in frequency over roughly the next thirty-percent of the graph, followed by a long 'tail' of low incidence motives which occur very few times. In all the tables of motives analysed in these corpora, this shape is relatively constant.

In conducting these analyses, incidence tables were not only collected from the entire corpora, but also from the individual expositions which make up the corpora. Comparison of these tables to those of the entire corpora demonstrate which works are most typical and atypical of the attributes seen in the corpora. The inclusion of all of these tables would not offer enough information to justify their space, but one example is included here, with interesting features highlighted. The tables below show the top ten highest incidence 4-tuples in the individual works of each corpus.[14] The percentages shown refer to percentage of material within each individual piece.

4-Tuple incidence table for works within the analysed corpus of Clementi											
Op. 1 No. 2 in G			**Op. 1 No. 4 in F**			**Op. 2 No. 4 in A**			**Op. 8 No. 3 in B♭**		
Motive	**Inc.**	**Perc.**	**Motive**	**Inc.**	**Perc.**	**Motive**	**Inc.**	**Perc.**	**Motive**	**Inc.**	**Perc.**
1, 1, 1	32	10.49	1, 1, 1	19	5.38	1, -1, -1	42	9.55	-1, -1, -1	23	10.36
-1, -1, 1	17	5.57	-1, -1, -1	14	3.97	-1, 1, -1	41	9.32	1, 1, 1	10	4.5
-1, -1, -1	15	4.92	-1, 1, -1	12	3.4	1, 1, 1	39	8.86	-1, -1, 1	8	3.6
-1, 1, -1	11	3.61	1, -2, 1	7	1.98	-1, -1, 1	38	8.64	7, -7, 7	5	2.25
-1, 1, 1	8	2.62	-2, 1, -2	7	1.98	-1, -1, -1	30	6.82	1, 1, -1	4	1.8
1, -1, -1	7	2.3	1, -1, 1	6	1.7	1, 1, -1	19	4.32	1, -1, -1	4	1.8
1, 1, -1	6	1.97	1, 1, -1	6	1.7	1, -1, 1	18	4.09	-1, 1, -3	4	1.8
0, -1, -1	5	1.64	0, 1, 1	5	1.42	-1, 1, 1	16	3.64	1, -3, 0	4	1.8
1, -1, 1	5	1.64	1, -1, -1	5	1.42	1, 1, -3	8	1.82	-3, 0, 0	4	1.8
-1, -1, 0	5	1.64	1, 1, -2	4	1.13	-3, 1, 1	7	1.59	0, 0, 1	4	1.8

14 Unlike the frequency tables for the entire sets of expositions, these lists are truncated to only ten entries. This is because the extended 'tail' section of the graph appears in higher rankings when the analysed corpora are smaller.

4-Tuple incidence table for works within the analysed corpus of Haydn											
No. 37 in E			No. 46 in E			No. 50 in D			No. 51 in B♭		
Motive	**Inc.**	**Perc.**	**Motive**	**Inc.**	**Perc.**	**Motive**	**Inc.**	**Perc.**	**Motive**	**Inc.**	**Perc.**
0, -1, 0	14	5	-1, -1, -1	90	29.03	-1, 1, 1	20	4.33	-1, -1, -1	48	12.77
-1, 0, -1	13	4.64	1, 1, -1	15	4.84	1, 1, -1	15	3.25	1, 1, 1	33	8.78
-1, -1, -1	12	4.29	-1, -1, 1	14	4.52	1, 1, 1	15	3.25	1, -1, -1	21	5.59
-1, -1, -1	9	3.21	-1, 1, 1	13	4.19	-1, -1, 1	14	3.03	-1, -1, 1	20	5.32
1, 1, -1	9	3.21	1, -1, -1	10	3.23	-2, -1, 1	13	2.81	-1, 1, -1	13	3.46
-1, 1, 1	6	2.14	1, 1, 1	7	2.26	-1, -1, -1	12	2.6	-1, 1, 1	12	3.19
0, -1, -1	5	1.79	3, -1, -1	6	1.94	1, -1, 1	11	2.38	0, 0, 0	11	2.93
1, -1, -1	4	1.43	-1, 1, -1	5	1.61	-1, 0, -1	10	2.16	1, 1, -1	9	2.39
-1, 1, 0	4	1.43	1, -1, 1	5	1.61	-1, 1, -1	10	2.16	3, 1, 1	5	1.33
1, 0, -1	4	1.43	-1, 3, -1	4	1.29	-1, 2, -1	9	1.95	-1, 2, -1	4	1.06

4-Tuple incidence table for works within the analysed corpus of Mozart											
No. 1 in C			No. 6 in D			No. 7 in C			No. 13 in B♭		
Motive	**Inc.**	**Perc.**	**Motive**	**Inc.**	**Perc.**	**Motive**	**Inc.**	**Perc.**	**Motive**	**Inc.**	**Perc.**
-1, -1, -1	69	16.31	-1, -1, -1	42	8.59	-1, -1, -1	39	10.4	-1, -1, -1	78	13.93
-1, -1, 1	16	3.78	1, 1, 1	17	3.48	1, 1, 1	25	6.67	1, 1, 1	32	5.71
2, -1, -1	13	3.07	-1, 2, -1	15	3.07	1, -1, -1	14	3.73	-1, 2, -1	25	4.46
-1, 2, -1	12	2.84	2, -1, -1	13	2.66	1, 1, -2	14	3.73	2, -1, 2	21	3.75
2, -1, 2	10	2.36	-1, 1, 1	11	2.25	-1, -1, 1	11	2.93	1, -1, -1	16	2.86
-1, 1, 2	10	2.36	1, 1, -1	11	2.25	-1, -1, 0	10	2.67	-1, 1, 1	15	2.68
2, 3, 2	10	2.36	-1, -1, 1	10	2.04	-1, 2, -1	9	2.4	-1, -1, 0	13	2.32
1, -1, 1	10	2.36	1, -1, -1	9	1.84	0, -1, -1	9	2.4	1, 1, -1	12	2.14
1, 2, -1	8	1.89	6, -6, 6	7	1.43	-2, -2, 1	7	1.87	-1, -1, 1	12	2.14
1, -1, -1	7	1.65	-5, 5, -5	7	1.43	-2, 1, 1	7	1.87	-1, 1, -1	11	1.96

These tables show some interesting contrasts between the summaries of the individual composer's corpora and the works within them. Some outliers which skew the corpus results are much clearer to see in this context — for instance, the fourth 4-tuple in the Op. 8 No. 3 Sonata by Clementi is a passage of alternating octaves. Octave figures occur surprisingly high in the Clementi tables in comparison to the Haydn and Mozart tables, but most of these are accounted for by only one work in the corpus. Similarly, Mozart's Sonata No. 6 features the disjunct (6, -6, 6) and (-5, 5, -5) passages in its ten highest ranked 4-tuples. While not high enough to make an appearance in the 4-tuple table for the entire corpus (those passages are ranked forty-fifth and forty-sixth), they still appear quite high in the rankings considering that they only occur in one bar in one exposition.

It is also curious to see that the Mozart corpus is the only one in which all four works have descending scalic figures as their highest ranked tuples, regardless of the results of the corpora as a whole. Descending scalic passages are the highest ranked tuples in the Haydn corpus, but not the highest ranked motive in each Haydn exposition — they reach this rank in the corpus as an average from the expositions.

For the most part, however, these single exposition tables reinforce the data found in the corpus tables. All of the scalic passages which occur frequently in the corpora are found within the most frequent material of the expositions. While an occasional section of figuration within an exposition may skew the results (such as in the Mozart Sonata No. 6 example), in general the highest ranked material in the expositions is the highest ranked material in the corpora.

3.2.2 - Interval Transition Matrices

All 3-tuple data in these corpora has been analysed, tallying the incidences of any interval being followed by any other interval. This data is shown in the tables on the following three pages, where each cell indicates the probability within each corpus (shown as a percentage) of the interval in the left-most column being followed by the interval in the top row.[15]

The transition values support what is generally found in the motivic incidence tables — for instance, since descending scalic motives feature so prominently in all composer's works, the transition probability from -1 to -1 is very high: 53.81% in Mozart; 51.47% in Haydn; and 46.11% for Clementi. These different values also reinforce the differences between corpora found in the incidence-tables, because Mozart's corpus features the highest ratio of descending scalic motives to other motives, and Clementi's corpus features the lowest. While the variations in statistical movement are significant, by colouring the cells which contain non-zero values (as is shown here), it is easy to see the variety of intervallic shapes each composer uses. As two examples of this, Mozart's corpus features a greater variety of possible intervals following an ascending fifth (ten against Haydn's seven and Clementi's five), while

15 A version of these tables which shows incidences, rather than percentages, is on the accompanying compact disc.

Clementi's corpus features a greater variety of possible intervals following an ascending octave (seventeen to Haydn and Mozart's seven).

As with the incidence tables, these matrices also display the many similarities these corpora share. The matrices all display clustering around their centre when the intervals are ordered by magnitude (ignoring relatively empty rows and columns produced by non-integer intervals). This provides further confirmation of the generally conjunct nature of writing in these corpora — transitions to disjunct intervals are generally always less likely than transitions to conjunct intervals. If the table is imagined as a Cartesian plane, where *x* and *y* axis values are provided by interval magnitude, clustering appears in quadrants I and III (this happens in all matrices but is especially visually noticeable in the Clementi corpus). This seems to suggest a general trend that melodic lines and figuration tend to change direction following a disjunct leap. This inference may be skewed by the amount of intervals in the corpus which follow tremolo-like patterns, where one interval is followed by the interval of the same distance in the opposite direction (as in figure 9). The occurrence of these patterns forms a diagonal line with a positive gradient; this is immediately noticeable in quadrants I and III of the Mozart corpus.

Before this experiment was conducted, it was postulated by the author that since these three composers would be well acquainted with the rules governing sixteenth-century counterpoint, that these rules may be apparent in interval transitions. Many of these rules are obviously inappropriate for this context (for instance, numerous intervals wider than a major-sixth occur melodically). However, one rule which these matrices allow to be tested empirically is described as follows:

> *An ascending leap of a minor sixth or an octave must be followed by a step back down within the compass of the leap. In the same way, a descending leap of an octave must be followed by a step back up within the compass of the leap* (McConnell).

This rule is often generalised to apply to any leap in either direction greater than a major or minor third (Cope, 2004, p. 13).[16] If this rule is followed by any of the composers, then there would be a clear tendency for disjunct intervals to be followed by a step in the opposite direction. Looking at the matrices, this hypothesis is supported to varying degrees. The Haydn corpus does not seem to favour opposite-direction-step transitions from any disjunct

16 Cope's paper actually goes so far as to disallow movement in the same direction following a leap of a third.

intervals — the highest percentage of transitions to a step from an opposite-direction leap is 31% (not including leaps greater than an octave). The Clementi corpus is similar, with one exception — leaps of an ascending seventh are followed by a descending second 63.64% of the time, with the other 36.36% of transitions accounted for entirely by inverted pedal-point figuration in Sonata Op. 8 No. 3 (6, -7 patterns). In contrast, 58.7% of ascending thirds and 36.17% of ascending sixths are followed by descending seconds in the Mozart corpus. It is unlikely that these differences are the result of the internalisation of any Fuxian aesthetic on the part of these composers, but they do expose some curious idiosyncrasies of the motivic vocabulary of each corpus. The agreement of these transition tables (and the other datasets) with Narmour's 'implication-realization' theory is discussed in section 3.3.1.

Interval Transition Matrix for the corpus of Wolfgang Amadeus Mozart

From \ To	-17	-16	-15	-14	-9	-8	-7	-6	-5	-4	-3.75	-3	-2	-1	-0.75	-0.25	0	0.25	0.75	1	1.25	1.75	2	3	3.75	4	5	6	7	8	9	10	14	16	18
-17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00
-16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	66.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	33.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	12.50	0.00	0.00	0.00	25.00	0.00	0.00	0.00	0.00	0.00	12.50	0.00	0.00	0.00	0.00	0.00	37.50	0.00	0.00	0.00	12.50	0.00	0.00
-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	23.53	0.00	0.00	5.88	0.00	0.00	5.88	0.00	58.82	5.88	0.00	0.00	0.00	0.00	0.00	0.00
-5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	28.00	0.00	0.00	16.00	0.00	0.00	0.00	44.00	12.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.55	0.00	9.09	0.00	0.00	4.55	0.00	0.00	9.09	0.00	0.00	13.64	4.55	0.00	40.91	9.09	4.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-3.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.13	0.00	0.00	31.91	2.13	0.00	0.00	2.13	0.00	0.00	21.28	0.00	0.00	10.64	2.13	0.00	8.51	6.38	2.13	10.64	0.00	0.00	0.00	0.00	0.00	0.00
-2	0.00	0.00	0.00	0.00	0.58	0.58	0.58	0.00	0.58	0.00	0.00	7.60	25.73	10.53	0.00	0.00	3.51	0.58	0.00	26.90	0.00	0.00	7.60	4.09	0.00	2.34	5.26	0.58	2.34	0.00	0.00	0.58	0.00	0.00	0.00
-1	0.00	0.00	0.00	0.15	0.00	0.30	0.00	0.00	0.46	0.30	0.00	0.91	3.96	53.81	0.00	0.30	6.71	0.15	0.00	15.09	0.00	0.00	11.43	0.76	0.00	1.37	2.29	0.30	1.07	0.30	0.00	0.15	0.00	0.00	0.15
-0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	0.00	0.00	0.00	0.00	0.00	0.00	1.69	0.85	0.00	0.00	0.00	0.85	10.17	35.59	0.00	0.00	18.64	1.69	0.00	13.56	0.00	0.00	7.63	3.39	0.00	1.69	2.54	0.00	0.85	0.85	0.00	0.00	0.00	0.00	0.00
0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	80.00	0.00	0.00	20.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.75	0.00	12.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	37.50	0.00	0.00	12.50	0.00	0.00	37.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.00	0.00	0.00	0.00	0.00	0.78	0.52	0.78	0.52	0.00	2.85	12.69	23.06	0.00	0.00	4.15	1.55	0.00	41.71	0.00	0.00	8.03	0.26	0.00	0.52	0.78	0.78	0.78	0.00	0.26	0.00	0.00	0.00	0.00
1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	50.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	1.09	0.00	0.00	0.00	0.00	0.00	2.17	1.09	0.00	1.09	6.52	58.70	0.00	0.00	2.17	0.00	0.00	7.07	0.00	0.00	10.87	9.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.70	2.70	0.00	16.22	8.11	21.62	0.00	0.00	8.11	0.00	0.00	5.41	0.00	0.00	35.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	87.50	0.00	0.00	12.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.45	25.81	3.23	3.23	12.90	19.35	0.00	0.00	12.90	0.00	0.00	9.68	0.00	0.00	3.23	3.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.51	17.02	6.38	0.00	4.26	2.13	36.17	0.00	0.00	4.26	0.00	0.00	19.15	0.00	0.00	0.00	0.00	0.00	0.00	2.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	9.52	38.10	14.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.76	0.00	0.00	33.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.33	0.00	0.00	0.00	16.67	16.67	16.67	0.00	0.00	29.17	0.00	0.00	4.17	0.00	0.00	8.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	33.33	0.00	0.00	33.33	0.00	0.00	33.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Interval Transition Matrix for the corpus of Joseph Haydn																															
From \ To	**-21**	**-11**	**-10**	**-8**	**-7**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**-0.25**	**0**	**0.25**	**0.75**	**1**	**1.25**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**21**
-21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.56	0.00	16.67	0.00	44.44	0.00	0.00	16.67	0.00	0.00	5.56	0.00	0.00	11.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.69	0.00	0.00	0.00	0.00	15.38	0.00	7.69	23.08	0.00	15.38	0.00	30.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.00	0.00	5.00	0.00	0.00	10.00	0.00	0.00	30.00	25.00	0.00	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.70	0.00	8.70	0.00	0.00	26.09	0.00	21.74	4.35	0.00	21.74	4.35	4.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	20.41	20.41	0.00	0.00	0.00	0.00	10.20	0.00	22.45	6.12	14.29	6.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-2	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	5.00	23.00	23.00	0.00	8.00	0.00	0.00	18.00	0.00	7.00	9.00	2.00	0.00	0.00	0.00	0.00	3.00	0.00	0.00	0.00	0.00	0.00	1.00
-1	0.00	0.20	0.00	0.00	0.39	0.20	0.20	0.20	0.98	2.35	51.47	0.20	9.00	0.20	0.00	22.31	0.00	4.70	2.15	1.57	1.37	1.17	0.39	0.00	0.20	0.20	0.20	0.00	0.20	0.20	0.00
-0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0	0.00	0.00	0.00	0.00	0.00	0.86	0.00	1.72	3.45	2.59	39.66	0.00	21.55	1.72	0.00	14.66	0.00	4.31	2.59	2.59	1.72	0.00	0.86	0.86	0.86	0.00	0.00	0.00	0.00	0.00	0.00
0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	66.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	33.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.00	0.00	0.00	2.21	1.26	1.58	1.26	1.89	6.94	28.71	0.00	2.52	0.00	0.00	41.96	0.00	4.10	2.52	1.26	2.21	0.00	1.26	0.00	0.32	0.00	0.00	0.00	0.00	0.00	0.00
1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	1.00	0.00	0.00	1.00	3.00	0.00	4.00	6.00	2.00	12.00	31.00	0.00	4.00	0.00	0.00	2.00	0.00	20.00	13.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	3.39	0.00	23.73	11.86	18.64	0.00	11.86	0.00	0.00	8.47	0.00	20.34	1.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	7.14	0.00	0.00	32.14	17.86	7.14	28.57	0.00	3.57	0.00	0.00	3.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	3.45	3.45	27.59	3.45	3.45	31.03	17.24	0.00	3.45	0.00	0.00	3.45	0.00	0.00	0.00	0.00	0.00	0.00	3.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	9.09	0.00	9.09	45.45	0.00	0.00	0.00	0.00	18.18	0.00	0.00	0.00	0.00	18.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	15.38	0.00	0.00	0.00	23.08	15.38	15.38	0.00	15.38	0.00	0.00	7.69	0.00	7.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	33.33	0.00	33.33	0.00	0.00	33.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Interval Transition Matrix for the corpus of Muzio Clementi																												
From \ To	-14	-8	-7.75	-7.25	-7	-6.75	-6.25	-6	-5	-4	-3	-2	-1	0	0.25	0.75	1	1.75	2	3	4	5	6	7	8	9	11	12
-14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	12.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	37.50	50.00	0.00	0.00	0.00
-7.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
-7.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
-7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.00	0.00	0.00	0.00	24.00	0.00	4.00	0.00	0.00	8.00	0.00	56.00	0.00	0.00	0.00	0.00
-6.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
-6.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
-6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	33.33	0.00	0.00	0.00	0.00	7.41	0.00	0.00	0.00	7.41	0.00	0.00	0.00	11.11	14.81	14.81	11.11	0.00	0.00	0.00	0.00
-5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	9.09	4.55	0.00	0.00	9.09	0.00	9.09	22.73	9.09	18.18	18.18	0.00	0.00	0.00	0.00	0.00
-4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.56	0.00	0.00	0.00	11.11	0.00	27.78	5.56	22.22	11.11	5.56	11.11	0.00	0.00	0.00	0.00
-3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	17.39	17.39	14.49	0.00	0.00	18.84	0.00	1.45	15.94	1.45	5.80	4.35	2.90	0.00	0.00	0.00	0.00
-2	0.00	0.00	0.00	0.00	2.30	0.00	0.00	0.00	1.15	1.15	11.49	17.24	10.34	11.49	0.00	0.00	27.59	0.00	8.05	1.15	1.15	6.90	0.00	0.00	0.00	0.00	0.00	0.00
-1	0.00	0.00	0.00	0.00	0.52	0.00	0.00	0.00	0.26	0.00	0.78	2.07	46.11	5.18	0.26	0.00	34.72	0.00	3.63	1.30	0.78	0.52	0.78	2.07	0.00	0.78	0.26	0.00
0	0.00	1.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.82	1.41	38.03	16.90	1.41	0.00	25.35	0.00	2.82	2.82	4.23	0.00	1.41	1.41	0.00	0.00	0.00	0.00
0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	66.67	0.00	33.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	33.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	66.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.26	0.00	0.00	0.00	1.04	0.00	0.00	0.26	1.04	0.78	6.48	5.44	29.02	1.30	0.26	0.00	44.30	0.00	3.89	2.33	0.52	1.30	1.04	0.52	0.00	0.00	0.00	0.26
1.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	1.64	0.00	0.00	0.00	1.64	0.00	0.00	0.00	0.00	4.92	1.64	26.23	32.79	11.48	0.00	0.00	3.28	0.00	9.84	6.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.50	10.00	0.00	40.00	2.50	12.50	10.00	0.00	0.00	17.50	0.00	5.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	35.00	20.00	15.00	20.00	0.00	0.00	0.00	10.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	13.79	0.00	0.00	0.00	0.00	0.00	0.00	37.93	0.00	17.24	13.79	0.00	0.00	0.00	0.00	6.90	0.00	10.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	36.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	63.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	7.50	2.50	2.50	17.50	5.00	2.50	15.00	2.50	10.00	5.00	7.50	5.00	5.00	0.00	2.50	5.00	0.00	2.50	0.00	2.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	66.67	33.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	50.00	50.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

These matrices also have the potential to generate short pitch-based motivic material in the style of any individual corpus. Any material longer than about 6 notes becomes functionally meaningless, because construction of themes and phrases requires more contextual information than mere averaged statistics (Meyer, 1957). While a random number generator may still produce some functionally nonsensical patterns by only picking the statistically insignificant outliers (consider a possible pattern in the Haydn corpus: 21, -3, -2, 21: a five note pattern spanning 5 octaves) for the sake of experiment and curiosity, several examples of different lengths are shown in figure 14 below:

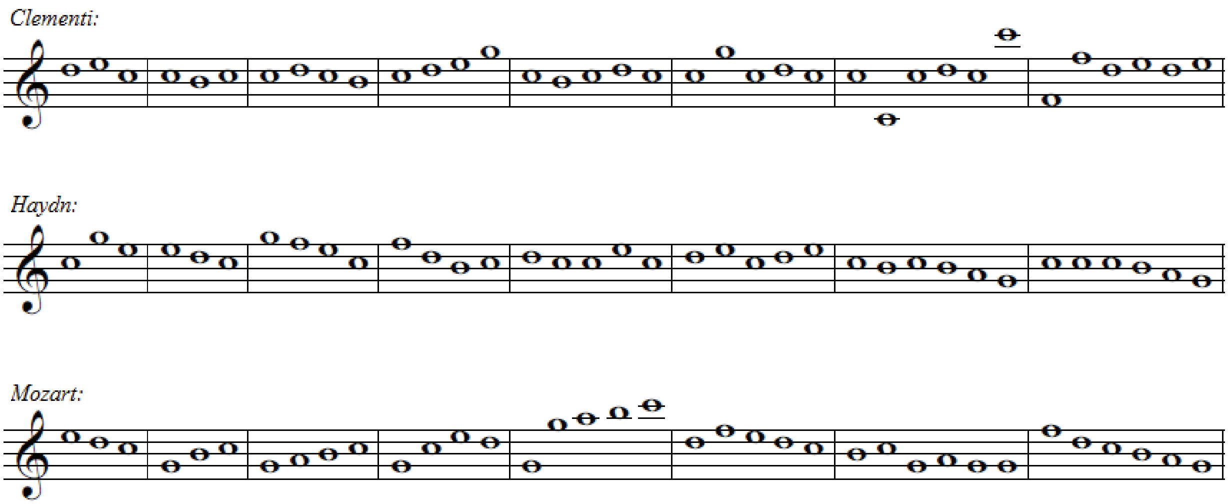


Fig. 14

These numbers were the first unique motives produced using random numbers from random.org (Haahr, 1998). Any numbers which produced chromatic intervals (non-integers) were re-generated, because chromatic intervals need to be in pairs to make sense under this system of notation (this only happened twice in the production of the above examples). These motives have been shown without any rhythm, but with possible harmonic implications in C major.[17] As predicted, the results contain a mixture of typical and atypical shapes, many of which could be expanded into themes and phrases typical of the corpora. For example, the 3-tuples, 4-tuples, and 5-tuples in the Clementi line outline chord shapes and common figuration, such as the turn (first 5-tuple). Some passages seem less typical of the figuration found in the corpora, however, such as the first Clementi 6-tuple, which is extremely disjunct. As the examples become longer, the potential for disorganisation increases, and this method of random construction becomes much less useful as a practical tool for musical composition.

17 It is important to remember that it is not these tonal patterns of scale degrees which are generated, but raw melodic contour — these patterns could start on any note, in any key.

3.2.3 - Measure of Disjunct Writing

The table below shows the disjunct writing score for all works and complete corpora.[18]

Muzio Clementi				
Op. 1 No. 2 in G	**Op. 1 No. 4 in F**	**Op. 2 No. 4 in A**	**Op. 8 No. 3 in B♭**	**Whole corpus**
0.706840	1.326761	0.531109	2.400670	1.099774
Joseph Haydn				
No. 37 in E	**No. 46 in E**	**No. 50 in D**	**No. 51 in E♭**	**Whole corpus**
0.918440	0.637821	1.193966	0.635582	0.872040
Wolfgang Amadeus Mozart				
No. 1 in C	**No. 6 in D**	**No. 7 in C**	**No. 13 in B♭**	**Whole corpus**
0.722941	1.351833	0.588196	0.706406	0.857008

Looking only at the disjunct writing scores for the whole corpora, it may be surprising that the Clementi corpus scores highest, despite the fact that the Clementi motive-incidence tables contain scalic material more prominently than either Mozart or Haydn. However, looking at the scores for individual works within the corpus reveals why this is so: the Op. 8, No. 3 sonata contains a much higher proportion of disjunct motion than the other works in the corpus. This is almost entirely due to the predominance of octave figuration in the right hand, and octaves appear much higher in this work's incidence table than in the rest of the corpus (7 is ranked third and -7 is ranked fifth in the 2-tuple table).

Given that a single work can heavily skew these figures for a whole corpus, it is more illuminating to examine the scores of single works. Doing this, we see that there is actually a surprising variation of disjunct writing between different pieces, but the averages between composers are not dramatically different. If the Op. 8 No. 3 sonata is removed from the Clementi corpus, the disjunct writing scores for each corpus are all between 0.8 and 0.9. The similar levels of variation in individual works suggests that perhaps each composer's use of disjunct or conjunct material can not be generalised over a corpus, but rather it is dynamic, and dependent on other compositional decisions made in each piece.

Despite this, curious trends can still be seen over entire corpora in the 2-tuple incidence tables. These function as tables showing the occurrence of all intervals within a corpus. Shown below are the entire 2-tuple tables for each corpus, ranked by incidence.

18 Note that the disjunct writing score for the whole corpus is not equal to the average of the four scores of the individual expositions — this is a similar, but different calculation.

Muzio Clementi			Joseph Haydn			Wolfgang Amadeus Mozart		
Motive	Incidence	Percent.	Motive	Incidence	Percent.	Motive	Incidence	Percent.
1	388	29.22	-1	511	35.58	-1	657	35.42
-1	387	29.14	1	318	22.14	1	387	20.86
-2	88	6.63	0	116	8.08	2	184	9.92
0	71	5.35	-2	102	7.1	-2	173	9.33
-3	69	5.2	2	100	6.96	0	118	6.36
2	61	4.59	3	59	4.11	-3	47	2.53
3	40	3.01	-3	49	3.41	5	47	2.53
7	40	3.01	4	29	2.02	3	37	1.99
5	29	2.18	5	29	2.02	4	31	1.67
-7	25	1.88	-4	23	1.6	-5	25	1.35
-5	22	1.66	-5	20	1.39	7	24	1.29
6	20	1.51	-7	18	1.25	-4	21	1.13
4	20	1.51	7	13	0.91	6	21	1.13
-4	18	1.36	-6	13	0.91	-6	17	0.92
-6	18	1.36	6	11	0.77	0.25	10	0.54
-8	8	0.6	9	6	0.42	0.75	8	0.43
8	4	0.3	0.25	3	0.21	-3.75	8	0.43
9	3	0.23	0.75	3	0.21	3.75	8	0.43
0.25	3	0.23	-8	1	0.07	-7	8	0.43
0.75	3	0.23	11	1	0.07	-8	3	0.16
-14	2	0.15	-11	1	0.07	8	3	0.16
11	2	0.15	13	1	0.07	-15	2	0.11
-6.75	2	0.15	14	1	0.07	-0.25	2	0.11
12	1	0.08	-10	1	0.07	10	2	0.11
1.75	1	0.08	8	1	0.07	1.75	2	0.11
-7.25	1	0.08	-21	1	0.07	-16	1	0.05
-7.75	1	0.08	12	1	0.07	-14	1	0.05
-6.25	1	0.08	21	1	0.07	1.25	1	0.05
			-0.25	1	0.07	18	1	0.05
			1.25	1	0.07	-17	1	0.05
			10	1	0.07	16	1	0.05
						9	1	0.05
						-9	1	0.05
						14	1	0.05
						-0.75	1	0.05

This table shows some surprising similarities. Both the Mozart and Haydn corpus have almost exactly the same proportion of their material accounted for by descending seconds — there is only 0.16% difference in this percentage between these two corpora. Looking further down, the first five most common intervals for both corpora are identical, even if they are in a different order. Clementi's table shows a curious counter-trend, but offers an even more startling fact: there are nearly exactly the same number of ascending seconds as descending seconds in the Clementi corpus (a difference of only 0.08%, or one interval). It is debatable whether or not these differences and similarities are significant features of the idiolect of each composer, but what is obvious for all three corpora is the limited number of intervals needed to construct most musical material. Using only unisons and ascending and descending seconds and thirds, 75.11% of the Clementi corpus, 79.86% of the Haydn corpus, and 81.89% of the Mozart corpus can be constructed.

The lower incidence intervals at the bottom of this table tend to indicate leaps which do not occur melodically, but rather indicate changes of register or incidences of widely voiced compound melody. If we acknowledge intervals larger than a ninth (8 or -8) to imply this shift in register, then we can quantify the occurrence of such passages in the corpus.[19] Under this criteria, the Clementi corpus features eight leaps from four unique intervals, the Haydn corpus features fifteen leaps from ten unique intervals, and the Mozart corpus features twelve leaps from ten unique intervals. This supports the conclusion that (at least within these corpora) Mozart and Haydn have a tendency to make a greater amount of registral leaps in their piano melodies and figuration than Clementi.

3.2.4 - Motivic Concentration

The following table shows the unique motives divided by total motives in each work and corpus for 3-tuples, 4-tuples, 5-tuples, and 6-tuples (the UML score).

19 Compound melodies can of course occur with much smaller intervallic separation, but smaller intervals occur melodically in addition to implying changes of register.

	3-Tuple UML	4-Tuple UML	5-Tuple UML	6-Tuple UML
Clementi Corpus	0.123867	0.296970	0.445289	0.558689
Cle. Op. 1 No. 2	0.212418	0.442623	0.598684	0.716172
Cle. Op. 1 No. 4	0.248588	0.498584	0.670455	0.777778
Cle. Op. 2 No. 4	0.158730	0.279545	0.382688	0.474886
Cle. Op. 8 No. 3	0.291480	0.423423	0.502263	0.545455
Haydn Corpus	0.115922	0.303221	0.459972	0.590141
Hay. No. 37	0.277580	0.503571	0.641577	0.744604
Hay. No. 46	0.196141	0.335484	0.469256	0.597403
Hay. No. 50	0.196544	0.389610	0.531453	0.623913
Hay. No. 51	0.204244	0.385638	0.512000	0.609626
Mozart Corpus	0.098325	0.246887	0.409116	0.548124
Moz. No. 1	0.162736	0.338061	0.495261	0.617577
Moz. No. 6	0.193878	0.388548	0.540984	0.655031
Moz. No. 7	0.151596	0.322667	0.489305	0.630027
Moz. No. 13	0.185383	0.364286	0.525939	0.655914

Since a lower UML score implies a higher motivic concentration in a corpus, this table implies that the Mozart corpus features the most reuse of pitch-based motives. This is far from a complete picture of a corpus's pattern of motivic reuse, however. Looking at the UML scores for the individual expositions, two points about the data are striking. Firstly, all individual exposition UML scores are higher than those of the corpora. This may seem counter-intuitive at first: it is reasonable to presume (from a musicological perspective) that there would probably be fewer unique motives in a single work than in a corpus of works, and that the addition of all the unique motives from the individual works would make the UML of the corpus higher. In actuality, many of the unique motives in the individual expositions are the same, so the total amount of unique motives in a corpus does not grow directly in proportion to the amount of pieces it contains. The ratio of unique motives to the total amount of material actually diminishes with each addition to the corpus, and this trend would continue with the addition of more works with similar musical vocabularies to those already used.

Secondly, there is a significant amount of variation in the UML scores between individual works. In the Mozart corpus, UML scores between works are significantly similar, generally not straying from the range of about 0.05. Haydn and Clementi both feature much more variation in scores between their individual works. Haydn's variation in UML scores occupies a range of around 0.08 to 0.18, Clementi's scores occupy a range of 0.14 to 0.3. It is interesting that this level of variation occurs despite the fact that these expositions were all chosen because of their similar musical specifications (as outlined in section 3.1). This perhaps suggests that any comparison between composers based on a limited corpus will produce easily misinterpretable results, and that a much larger selection of works may be needed to make inferences about idiomatic motivic reuse.

Another interesting application of this data is in showing how the length of motives under examination factor into varying motivic concentration scores between pieces and corpora. Since there are more possible permutative possibilities as *n* increases, it is logical that UML scores will increase as they operate on longer length *n*-tuples. This is reflected in the table. However, these results do not increase equally for all works, as some works will have a predisposition towards motivic 'chunks' of certain lengths (as mentioned in section 2.3.4). This can be seen, for example, in the scores of Haydn's No. 50 and No. 51 expositions. No. 51 has a higher 3-tuple UML score than No. 50, but the 4-tuple, 5-tuple, and 6-tuple scores are all higher in No. 50. This suggests, even if only slightly, that the No. 50 exposition is better able to account for repeated 4, 5, and 6 pitch figures, yet No. 51 better accounts for the repetition of 3 pitch figures. Looking at the results between the corpora, the only overlap in scaling occurs between the Haydn and Clementi corpora. It may be tentatively postulated that the Clementi corpus better accounts for 3-tuples, while there is more concentration of longer *n*-tuple figures in the Haydn corpus.

The table below shows another set of results — various MC*n* scores, with different values of *n* used for 3-tuples, 4-tuples, 5-tuples, and 6-tuples. The values of *n* have been chosen to focus on the top of each incidence-table because that is where there is the most interesting variation in incidences between each motive.[20] In order to examine a similar quantity of motives from each motive bank, the value of *n* must decrease as the length of the motives being examined becomes longer. This is because in the longer motive-incidence tables there are more possible combinations, and musical material tends to be spread between a greater quantity of unique motives.

20 Also, the inclusion of any attempt at an exhaustive list of MC*n* values for various n-tuples would be extremely space-consuming, and largely, not illuminating.

	3-Tuple MC50%	4-Tuple MC30%	5-Tuple MC20%	6-Tuple MC15%
Clementi Corpus	7	6	8	9
Cle. Op. 1 No. 2	7	7	9	13
Cle. Op. 1 No. 4	12	16	17	19
Cle. Op. 2 No. 4	3	4	4	3
Cle. Op. 8 No. 3	11	10	9	7
Haydn Corpus	7	7	8	12
Hay. No. 37	10	11	10	10
Hay. No. 46	3	2	1	1
Hay. No. 50	12	12	12	12
Hay. No. 51	4	4	5	6
Mozart Corpus	7	8	9	13
Moz. No. 1	6	6	6	7
Moz. No. 6	11	11	11	11
Moz. No. 7	6	6	9	12
Moz. No. 13	6	5	5	6

Here again, there are similar figures in the results for all the corpora, with a much greater level of variation between single works. This seems to reinforce the conclusion from the previous table — that the similarity of results between corpora, considering the differences in the expositions, is probably co-incidental. A larger corpus of works is needed to investigate this. The most interesting entries in this table are the outliers. Mozart's Sonata No. 6, Haydn's Sonata No. 50, and Clementi's Op. 1 Sonata No. 4 show examples of low motivic concentration, where many motives are needed to account for the musical material. This suggests a high level of pitch motive variety in the melodic contour of work. Conversely, Clementi's Op. 2 Sonata No. 4 and Haydn's Sonata No. 46 show examples of high motivic concentration, where a large amount of musical material is accounted for by a smaller amount of unique motives. This suggests a more limited amount of variety in the melodic contour of the work. Looking at the score, this is exactly what we see — these two sonatas are mostly scalic and filled with repeated figures which easily reduce to recurring motives.

While general agreement between the trends of the UML and MC*n* tables is expected, it is interesting to see that the disjunct writing scores also follow the trends shown above. This is easy to understand — an excerpt which features a large amount of scalic writing (having a low disjunct writing score) has a large amount of material which can be accounted for in a few scalic motives. This causes low UML and MCn scores, and hence implies a high level of motivic concentration.

The UML and MCn scores, while providing a general summary of motive distribution, do not provide a complete explanation. They can still be manipulated to provide misleading results. For a full understanding of motive distribution in a corpus, every motive and its incidence needs to be shown. This can be done graphically. The following four graphs show the motivic incidences for the three corpora.[21] The x-axis on each (motive rank) is shown logarithmically, to provide detail of the variation at the top of the incidence tables, and to compress the large quantity of motives following this. Figure 15 shows the 3-tuples, figure 16 shows the 4-tuples, figure 17 shows the 5-tuples, and figure 18 shows the 6-tuples. In each graph, the Clementi corpus n-tuples are shown in red, Haydn corpus n-tuples in blue, and Mozart corpus n-tuples in green.

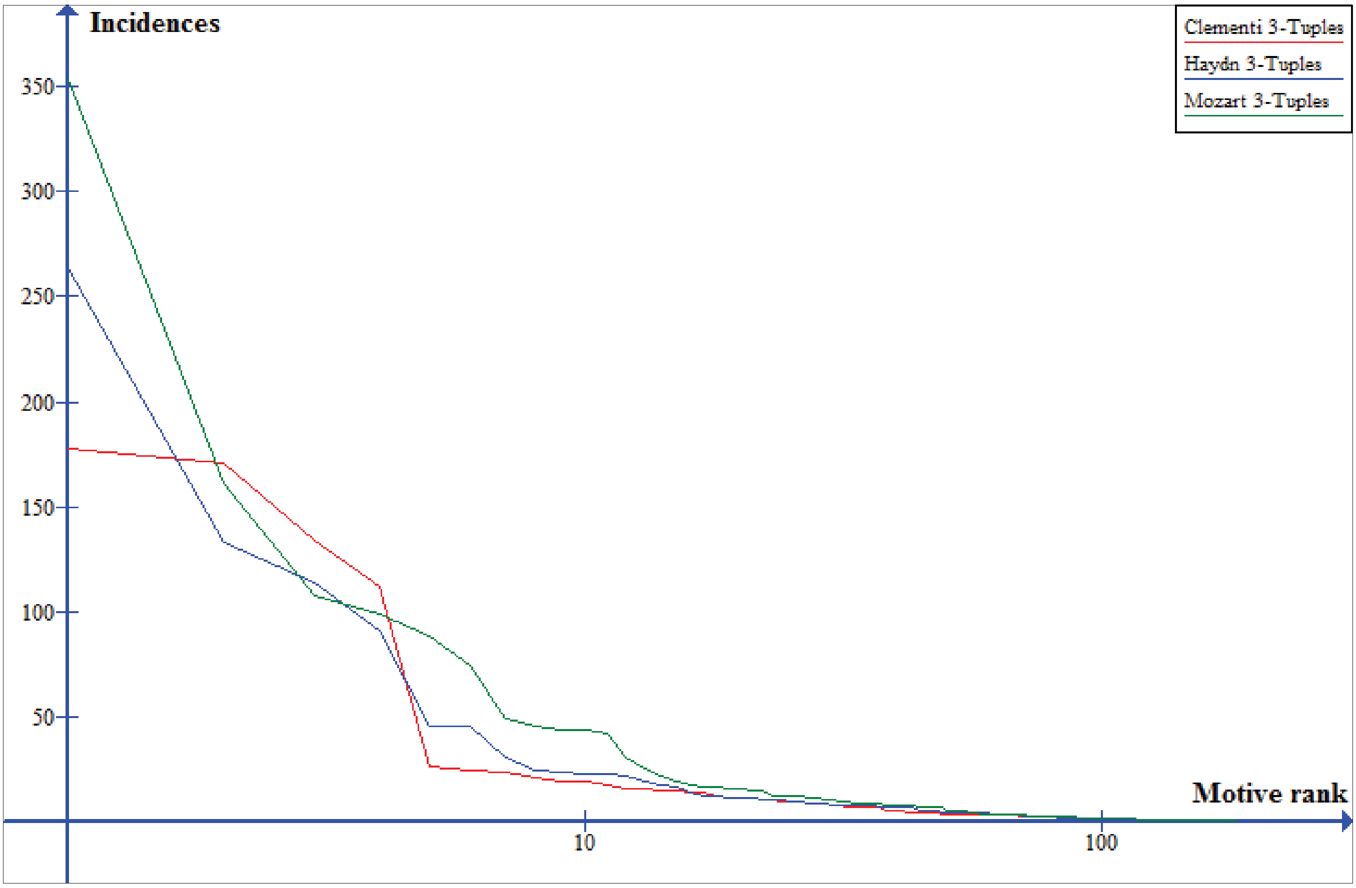


Fig. 15

21 While this set of incidence and rank co-ordinates is obviously a set of points which is non-contiguous, it is shown in these graphs as a line for clarity.

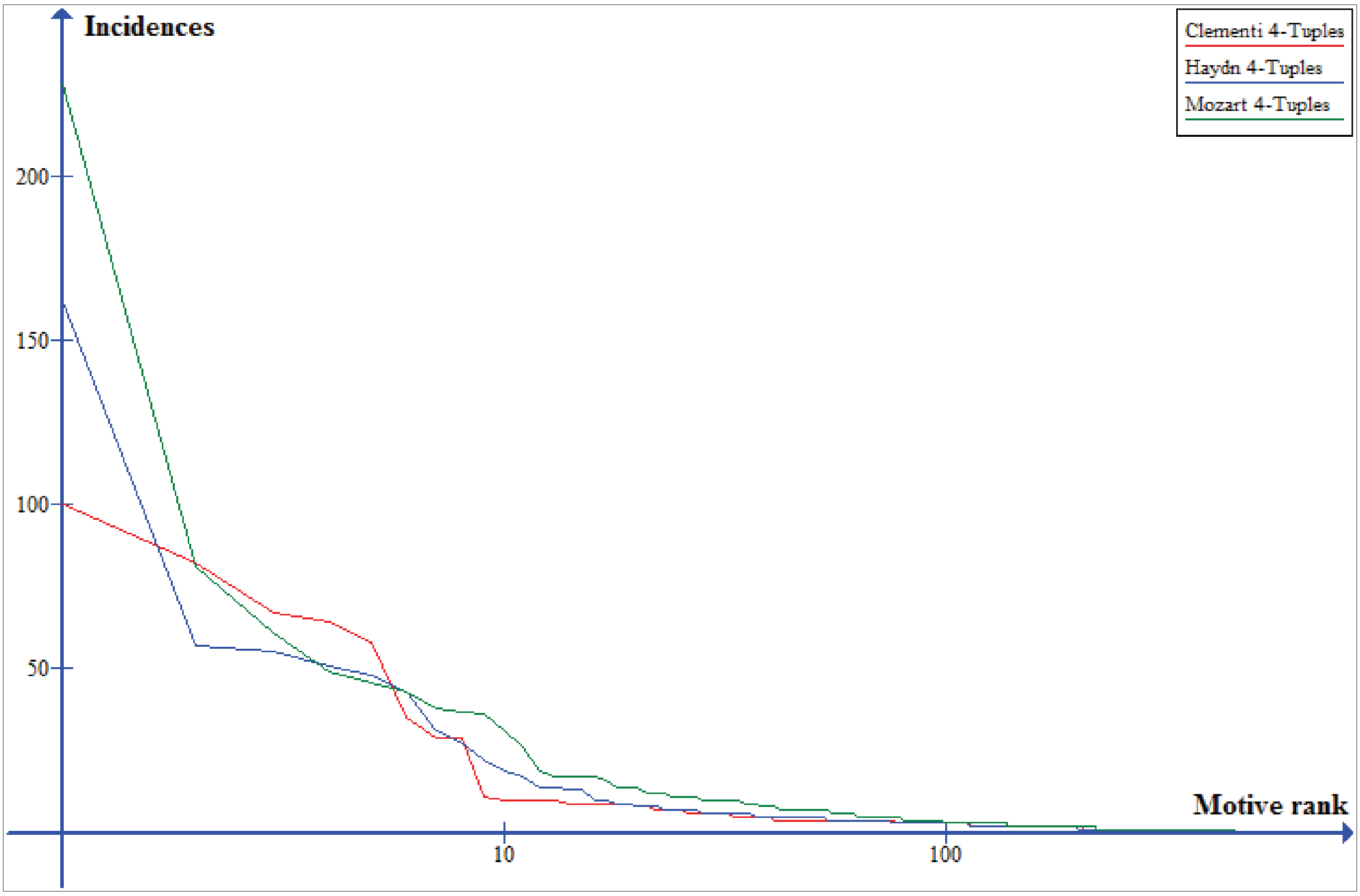

Incidences
Clementi 4-Tuples
Haydn 4-Tuples
Mozart 4-Tuples
200
150
100
50
Motive rank
10
100

Fig. 16

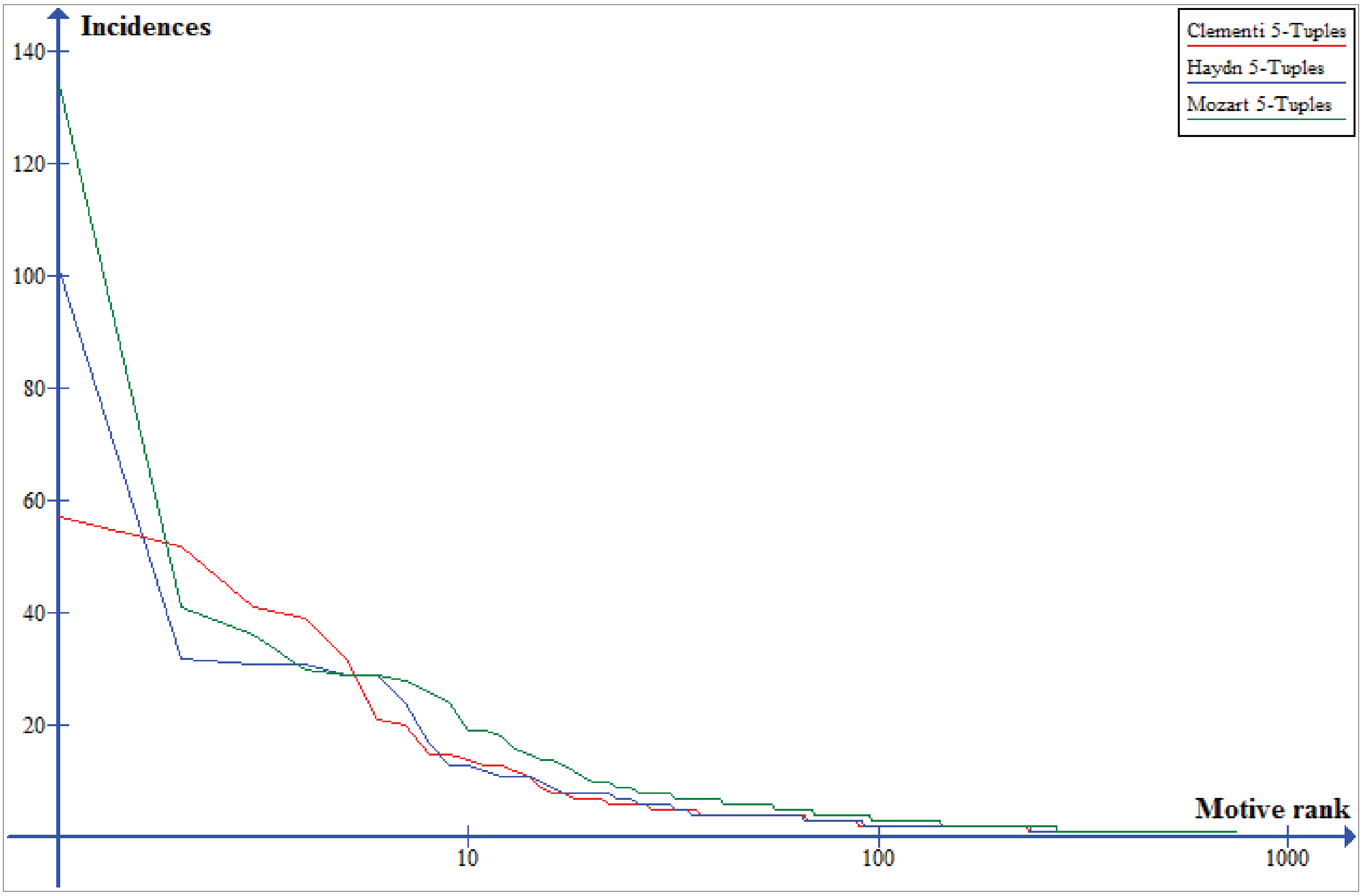

Incidences
Clementi 5-Tuples
Haydn 5-Tuples
Mozart 5-Tuples
140
120
100
80
60
40
20
Motive rank
10
100
1000

Fig. 17

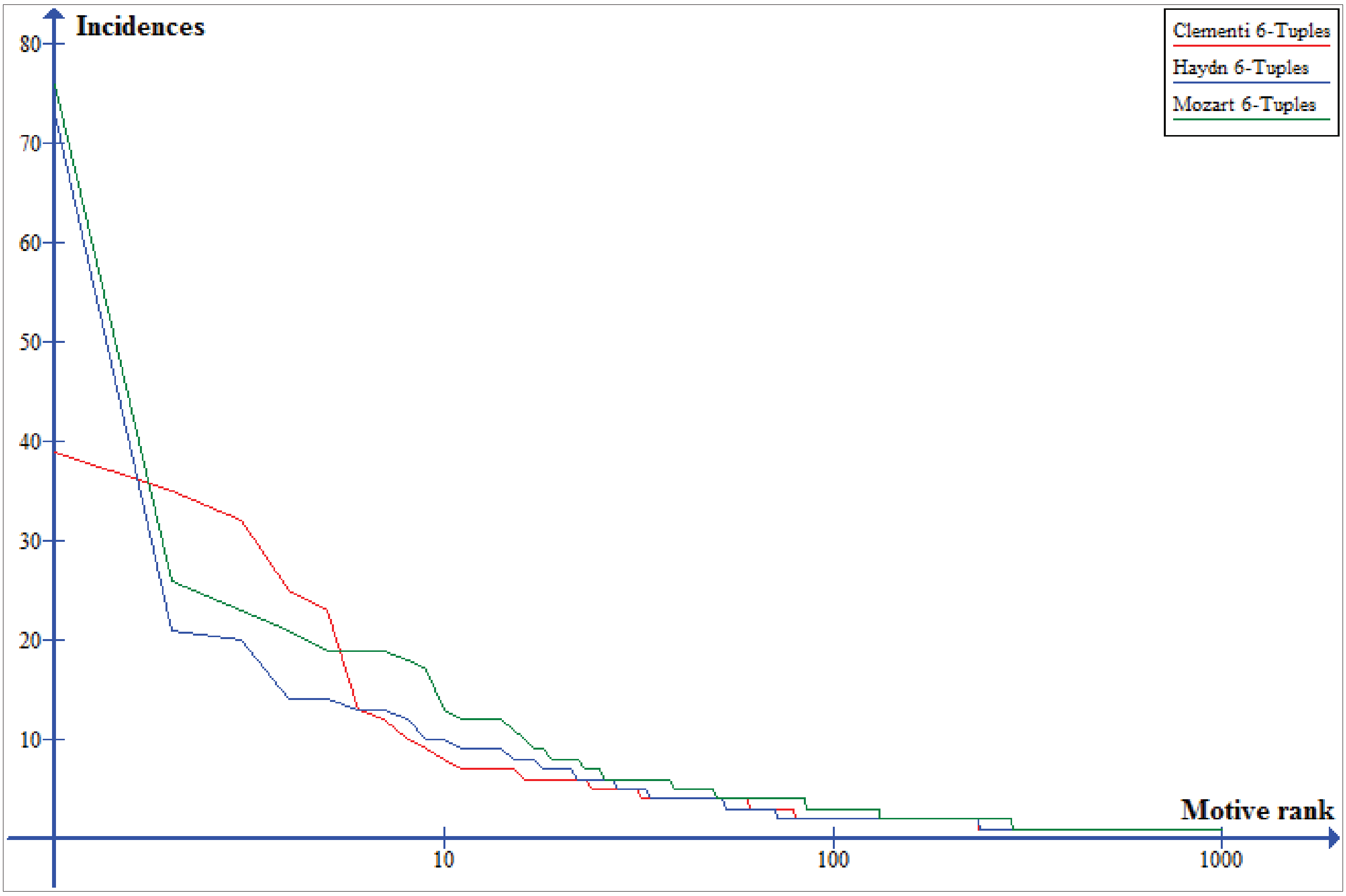


Fig. 18

These graphs all show that while the initial few entries of each incidence table may be quite different, the distributions all start to converge after around the twentieth rank. It is interesting that the results for the corpora of the three composers all converge to the same shape, despite them consisting of different amounts of material — there are more motives (both unique and repeated) in the Mozart corpus than the Haydn or Clementi corpus, yet all three lines follow an extremely similar distribution as incidences become lower. This suggests that regardless of the sample sizes used in this study, there is a similar logarithmic distribution which occurs, with a steep drop off of incidences after the first 10-20 motives. It is hypothesised from these results that increasing the amount of works in the corpora (or the length of material that is examined) would maintain the shape of the above distributions, while simply increasing the magnitude of the incidences and ranks to be shown.

As mentioned in 3.2.1, the distribution of all of these rank\incidence co-ordinates follows a logarithmic distribution. As we move from 3-tuples to 6-tuples, this distribution stays similar, but the *y*-axis is compressed, and the *x*-axis is stretched. Since the results for individual expositions also follow this general distribution, it is probable that for any genre of music which relies upon a significant degree of pitch contour repetition (classical-era music is but one example of many) a logarithmic distribution of motive incidences will emerge if this system is used.

3.3 - Discussion

These results probably do not, by themselves, say anything profound about the differences between these corpora. On the whole, the Clementi corpus is the most disjunct, and the Mozart corpus frequently features disjunct shapes more prominently than the other two corpora. Also, the Clementi corpus features ascending and descending scalic motion in equal amounts, unlike the Haydn and Mozart corpora, where descending scalic passages dominate. Certainly, these facts could be seen as significant conclusions, but many of the most interesting differences occur between works in a single corpus. The large amount of variation found between these individual expositions in sections 3.2.3 and 3.2.4 suggests that these composers have a variable pitch-based vocabulary which is utilised differently in different pieces. Even in these sonata expositions, which were selected to be similar, there are very different sets of motives in the incidence tables for each work.

Possibly more than highlighting the trivial differences between each corpus, this research highlights many striking similarities. The highest ranking entries in the *n*-tuple frequency tables between all the pieces and composers show many shared motives. This indicates that much of the common figuration in each piece draws from a set of typical classical-era shapes and patterns. The extremely high reuse of scalic passages in the corpora is also shown to be common, by the most frequent intervals and the highest motives of every *n*-tuple table. Also, the graphs in section 3.2.4 suggest that the distribution of motive use in all of these corpora follows a typical logarithmic curve. While these conclusions are only supported by limited sampling from the classical repertoire, perhaps the most important idea to take away from this research is that these results are incapable of proving or disproving anything profound about the nature of composition by themselves. There may be other attributes of these pieces which are not examined here which reveal shocking revelations when compared between composers. However, the data generated so far seems to suggest more than anything else that the differences in melodic figuration between these corpora are trifling and insignificant in comparison to their similarities.

3.3.1 - Concordance with Existing Research

While occasional references to pre-existing informal notions or specific research have been made in previous sections, some conclusions from existing research are supported by multiple applications of the results, and hence merit discussion in a separate section.

David Huron's research into the melodic contour of folk song corpora reveals many of the same conclusions as the data shown above. In particular, the fact that descending scalic passages account for so much material in the pieces in this study seems to be something which is at least a generalisation of Western tonal music. In Huron's study of 'The Melodic Arch in Western Folksongs', he generalises phrase motion in his conclusion; “What goes up is likely to come down, but what goes down is less likely to come back up”. Additionally, he summarises intervallic tendencies as generally conforming to a pattern of ascent by leap, and descent by step in *Sweet Anticipation*. These tendencies are expressed by the overwhelming tendency towards descending scalic figures in the *n*-tuple tables, and the scarcity of ascending skip-ascending skip transitions in the interval transition matrices.

Additionally, both von Hippel and Huron make reference to the following table, which describes all stepwise movement probabilities over a set of two intervals for a large collection of Western and non-Western works:

Percentages of Transitions which Involve only Stepwise Intervals			
Descending followed by Descending	70%	Descending followed by Ascending	30%
Ascending followed by Descending	51%	Ascending followed by Ascending	49%

These intervallic transitions are very similar to the findings in this study. The following table expresses the above statistical combinations as they are found in the Haydn, Mozart, and Clementi corpora (extracted from 3-tuple tables, where D stands for Descending and A for Ascending):

Percentages of Transitions which Involve only Stepwise Intervals in each Corpus											
Clementi				Haydn				Mozart			
D, D	57%	D, A	43%	D, D	70%	D, A	30%	D, D	78%	D, A	22%
A, D	40%	A, A	60%	A, D	41%	A, A	59%	A, D	36%	A, A	64%
Percentages of Stepwise Transitions Combining the Three Corpora Above											
D, D			70%			D, A			30%		
A, D			39%			A, A			61%		

These results agree strongly with the findings of von Hippel and Huron. It is particularly striking that when the corpora are combined, the proportions of descent-descent to descent-ascent transitions are within a percent of the same results operating on a much wider corpus of works. Clementi's corpus possesses the weakest propensity towards descending scale rather than ascending scale figures, as shown in the *n*-tuple lists, and this is reflected in these percentages. The increased bias towards ascending scalic rather than ascent-descent figures (in comparison to the almost equal split in the general corpus) may well be a feature of the particular style of music undergoing analysis here. For this information to be conclusive requires a much larger analysed corpora.

It is also worth examining to what extent the intervallic transitions of this study satisfy the innate principles of Narmour's 'implication-realization' theory. The way this theory explains the majority of interval transitions can be summarised as follows (small intervals are defined as up to 5 semitones, large intervals are 7 or more, and tritones can go both either way depending on their spelling)[22]:

1. Small intervals imply a following small interval in the same direction.
2. Large intervals imply a change in direction by the following interval.
3. Small intervals imply a following similarly sized interval.
4. Large intervals imply a following smaller interval.

Narmour's system can not be perfectly implemented with the figures of this analysis, as it relies on knowledge of chromatic interval states, and in the research in this thesis intervals of

22 These four points only summarise the Principle of Registral Direction (PRD) and the Principle of Intervallic Difference (PID) from Narmour's system. The 'implication-realization' theory actually covers much more than this, with separate coverage of what devices constitute musical closure, and a thorough discussion of the role of learned musical associations. Additionally, there are other cases not covered here, such as the perception of some altered processes retrospectively. However, the above four cases are the most frequently cited and tested principles from the theory, and account for a massive amount of musical material.

different chromatic proportions can be treated as identical (depending on context). However, it is still possible to quantify the general proportion of interval transitions that are covered by these principles. The second and fourth principle above describe typical Fuxian counterpoint writing conventions, so the observations made in support of these conventions in 3.2.2 also apply to Narmour's theory. Looking again at the table of the 30 most frequently occurring 3-tuples, there is only one motive which does not conform to these tendency patterns — the (-1, 7) interval which occurs 8 times in the Clementi corpus. Many more exceptions to the above principles exist, but don't show up on the tables because they occur so infrequently. This supports what intuition would suggest — the vast majority of material here is accounted for by predictable intervallic transitions, with a comparitively miniscule amount of less predictable transitions. These unpredictable transitions may occur between phrases, signifying registral shifts, or they may simply occur as novel melodic figures.

4 - Further Applications

While the system as outlined above could find further use as a means of corpus-based melodic comparison, there is also the potential for different applications in the future.

4.1 - Identifying Compositional Idiolect

As used so far, the programs for collecting and comparing *n*-tuple data have only operated on very small corpora. If these corpora were expanded to be more representative of a composer's output as a whole, then it may be possible to establish some more firm conclusions and make more definitive claims. For example, it would be possible to say which shapes are the most prominent in all the melodic writing of Mozart. Or perhaps it could be seen if the distribution patterns seen in section 3.2.4 are typical of the compositional idiolect of each composer in this study, rather than just a select subset of their works.

With a larger set of works from which to draw conclusions, the specialised measurements of motivic idiom used in this study (such as the UML and MC*n* scores), may become more significant if their value is reinforced by hundreds of works. It will be fascinating if trends emerge which are consistent over a large amount of analysed works. It will be especially significant if these trends are consistent within corpora from single composers, but vary between different composers. If these outcomes occur when the entire output of the composers used in this study is examined, this would suggest that composers of the classical-era do tend to reuse motives in significantly different ways from each other. Classical-era music provides only one practical area of examination, however — the analytical methods outlined in this thesis may be used on any composer from any era to make judgements about their melodic motivic idiolect.

To carry out analysis on a composer's entire compositional output would require that many scores are reduced to melodic lines, and these melodic lines then converted to the numeric format used here. In this thesis, for relatively small corpora, this was done by hand by a single researcher. For larger corpora, either a team or automation is necessary. The latter option is discussed in section 4.2.

4.1.1 - Jazz Improvisation

The comparison of motivic use in jazz improvisation is another area of possible exploration outside traditional written scores. While less useful in modern ‘post-tonal’ genres of jazz (such as ‘free jazz’), in any genre of jazz where a recorded artist has performed a solo over an established set of chord changes, meaningful comparisons regarding motivic idiolect can be made. It is relatively easy to discuss an artist's use of tonality — this can be done either aurally or through analysing transcriptions of solos. It is less easy to make more generalised statements about which particular motives occur most frequently in an artist's solos, or the extent to which the patterns which make up these solos are reused.

It is widely acknowledged that jazz musicians rely on permutations of enclosures and scale and chord shapes as the fundamental building blocks of their vocabularies in tonal jazz (Nelson, 1966). Such a comparison is far outside the scope of this research, but in practice it would be possible to compare motivic use in the recorded solos of Charlie Parker against the recorded solos of Dizzy Gillespie, for example. What results this might produce cannot even be conjectured, but the comparison would no doubt be interesting. Potentially, with enough researchers, or enough time, comparison could be made across a large group of artists. As an example, it would be possible to see the most frequent shapes employed by all of the leading Bebop alto saxophonists of the 1950s. Such answers may have implications for future pedagogy, and may even assist in tracing historical stylistic influences.

For this to be a valid method of analysis, care must be taken with the transcription to account for enharmonic ambiguities which arise in extended harmonies. But even if harmonic ambiguities are included in the transcriptions, if shapes are uniformly presented, the current method of representation is perfectly suited for the tonal fragments which are employed in most jazz improvisation.

4.1.2 - Composer Identification

Much as corpus-based algorithms are used to differentiate composers for the field of music information retrieval, predictions about the origin of works of unknown authorship could be made with the systems of comparison outlined in this thesis. This may be difficult if the level of variation shown in sections 3.2.3 and 3.2.4 proves to be a general trend for all of a typical composer's output. It still may be possible to make meaningful matches, however, if a work is compared with a subsection of a composer's corpus which has similar features as the work being identified.

If such a system of classification were even possible, this might change the conclusions reached in section 3.3 — when taken all together, the values generated by this system of analysis could provide a unique ‘fingerprint’ of a composer's motivic reuse. This seems unlikely, however, given that even relatively small changes in a composer's style may produce works which are more typical of a different composer's corpus. It is more likely that the general level of variation present makes the classification of works by composers of similar styles extremely difficult (or even impossible) with this system.

If the works being identified fall into very different genre classifications, however, the situation may be quite different. In this case, significant differences in motivic idiolect probably do exist, making automated classification by this system possible.

4.1.3 - Limitations in Chromatic Writing

If this system is used in the analysis of Western music written after the classical era, problems are encountered as chromaticism becomes a distinctly motivic feature in a manner different to that found in classical-era works. Any works which use symmetrical divisions of the octave as a tonal device will have elements which are fundamentally inexpressable with this system, because it is based around traditional tonal key structures. Notes in tonal systems which use enharmonic spellings interchangeably will not be properly motivically represented. For these to be represented properly, the system would need to be adapted to function chromatically, and for motives to be stated in chromatic, rather than diatonic steps. Potentially, this modification (which would only affect the nature of the notation, not the operations involved) would then allow this method to be used on a corpus of atonal melodies. Obviously, for some serialist structures, this may mean a combinatorial explosion of single-occurrence motives, and no new useful information gained. But in some cases, it may highlight repeated patterns which are otherwise difficult to perceive within a corpus.

There is, however, a large body of works of modern idiom which could still be analysed by the system as is. Works which use quartal or quintal harmony (such as those by Hindemith) tend to preserve motivic relations on the staff — for instance, a chord comprised of layered fifths will generally have this spelling of fifths preserved wherever it appears on the staff. A chord made out of stacked tritones, however, may change if the spelling of a diminished fourth and augmented fifth are interchangeable. Also, works which utilise seven-note modes (such as many folk songs presented by Bartók, Kodály, and others) will still use a tonal structure which is compatible with this system. There is also nothing stopping this system from being used with any purely modal music using modes with six or fewer notes, provided the notation is selected so that the modes appear with the same spacing on the staff in each piece.

4.2 - Integration with Existing Software

For this thesis, numeric reduction was done by hand from the original piano scores. This is a very time-consuming process, and it is not practical when using large corpora. If the process was automated, the analysis of long or numerous works becomes trivial. This is feasible, by incorporation of the system of numeric reduction into music engraving software such as Finale or Sibelius. Many classical scores exist in files which are readable by this software. The reduction to a single melodic line can be done either perceptually using human analysis, or with heuristic devices from other research. Once this is completed, all that remains is to convert a single melodic line into the numerical format.

The process of numeric reduction is almost entirely algorithmic, featuring only rare moments where a musicologist must exercise judgement about how to represent a passage. These moments of subjective judgement can be handled by chromatic heuristics, which can have adjustable thresholds for dealing with the representation of chromatic passages. This would be easily implementable as a plug in, which could automatically output a text file which is ready to be entered into the motive programs being used. Inbuilt scripting languages, such as *ManuScript* (Sibelius), or *FinaleScript* (Finale) would allow the relatively easy implementation of melodic extraction algorithms or data format encoding. An intuitive text based system of score encoding such as that used by Lilypad would make encoding operations especially easy.

4.3 - Closing Comments

The systems presented in this research provide a new and unique means of comparison between melodic corpora. While there are acknowledged limitations in the data representation, it is my hope to see a system such as this incorporated into future analytical musicology. If this research does nothing more than inspire other scholars with the future possibilities of statistical analysis, this alone is a highly positive outcome. The incorporation of scientific quantification into traditional areas of musical analysis is still a field which is thoroughly worthy of fresh exploration.

5 - Bibliography

Abdallah, S., Noland, K., Sandler, M., Casey, M., & Rhodes, C. (2005). 'Theory and Evaluation of a Bayesian Music Structure Extractor.' *Paper presented at the 2005 International Society for Music Information Retrieval Conference.* London.

Adams, N. H., Bartsch, M. A., Shifrin, J. B., & Wakefield, G. H. (2004). 'Time Series Alignment for Music Information Retrieval.' *Paper presented at the 2004 International Society for Music Information Retrieval Conference.* Barcelona.

Alphonce, B. H. (1980). 'Music Analysis by Computer: A Field for Theory Formation.' *Computer Music Journal,* 4(2).

Barthélemy, J., & Bonardi, A. (2001). 'Figured Bass and Tonality Recognition.' *Paper presented at the 2001 International Society for Music Information Retrieval Conference.* Bloomington, Indiana.

Bellman, H. G. (2011). 'Categorization of Tonal Music Style - A quantitative investigation.' Doctoral Thesis from Queensland Conservatorium.

Beran, J., & Mazzola, G. (1999). 'Analyzing Musical Structure and Performance: A Statistical Approach.' *Statistical Science,* 14(1).

Blood, B. (2011). 'Music Theory Online: Ornamentation.' Retrieved November 16th, 2011, from www.dolmetsch.com/musictheory23.htm

Bod, R. (2002). 'A Unified Model of Structural Organization in Language and Music.' *Journal of Artificial Intelligence Research,* 17.

Brinkman, A. R. (1986). 'Representing Musical Scores for Computer Analysis.' *Journal of Music Theory,* 30(2).

Brown, A. R. (2004). 'An aesthetic comparison of rule-based and genetic algorithms for generating melodies.' *Organised Sound,* 9(2).

Burkholder, J. P., Grout, D. J., & Palisca, C. V. (2010). *A History of Western Music* (8th ed.). New York: W. W. Norton & Company.

Cambouropoulos, E. (2000). 'Melodic Clustering: Motivic Analysis of Schumann's Träumerei.' *Paper presented at the III Journées d' Informatique Musicale.* Bordeaux, France.

Cambouropoulos, E., & Widmer, G. (2000). 'Melodic Cue Abstraction, Similarity and Category Formation: A Computational Approach.' *Paper presented at the 2000 International Conference of Music Perception and Cognition.* Staffordshire, England.

Chai, W. (2005). 'Detection of Key Change in Classical Piano Music.' *Paper presented at the 2005 International Society for Music Information Retrieval Conference.* London.

Conklin, D. (2003). 'Music Generation from Statistical Models.' *Paper presented at the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences.* Aberystwyth, Wales.

Cope, D. (2001). *Virtual Music: Computer Synthesis of Musical Style*. Cambridge, Massachusetts: MIT Press.

Cope, D. (2002). 'Computer Analysis and Composition using Atonal Voice-Leading Techniques.' *Perspectives of New Music,* 40(1).

Cope, D. (2004). 'A Musical Learning Algorithm.' *Computer Music Journal,* 28(3).

Deutsch, D. (1991). 'Pitch Proximity in the Grouping of Simultaneous Tones.' *Music Perception,* 9(2).

Dubnov, S., Assayag, G., Lartillot, O., & Bejerano, G. (2003). 'Using machine-learning methods for musical style modeling.' *Computer,* 36(10).

Fry, C. (1980). 'Computer Improvisation.' *Computer Music Journal,* 4(3).

Gjerdingen, R. O. (2007). *Music in the Galant Style*. Oxford: Oxford University Press.

Grachten, M., Arcos, J. L., & Mántaras, R. L. d. (2005). 'Melody Retrieval using the Implication/Realization Model.' *Paper presented at the The International Conference of Music Information Retrieval.* London.

Haahr, M. (1998). 'RANDOM.ORG.' Retrieved December 2nd, 2011, from http://www.random.org/

Hiller, L., & Bean, C. (1966). 'Information Theory Analyses of Four Sonata Expositions.' *Journal of Music Theory,* 10(1).

Hiller, L., & Ramon, F. (1967). 'Structure and Information in Webern's Symphonie, Op. 21.' *Journal of Music Theory,* 11(1).

Hippel, P. v., & Huron, D. (2000). 'Why Do Skips Precede Reversals? The Effect of Tessitura on Melodic Structure.' *Music Perception,* 18(1).

Hoffmann, E. T. W. (1989). 'Review of Beethoven's Fifth Symphony' (M. Clarke, Trans.). In D. Charlton (Ed.), *E.T.A. Hoffmann's Musical Writings: Kreisleriana, The Poet and the Composer, Music Criticism* (pp. 234-257). Cambridge: Cambridge University Press.

Hörnel, D., & Menzel, W. (1998). 'Learning Musical Structure and Style with Neural Networks.' *Computer Music Journal,* 22(4).

Huron, D. (1996). 'The Melodic Arch in Western Folksongs.' *Computing in Musicology,* 10.

Huron, D. (2001). 'Tone and Voice: A Derivation of the Rules of Voice Leading from Perceptual Principles.' *Music Perception,* 19(1).

Huron, D. (2006). *Sweet Anticipation*. Cambridge, Massachusetts: MIT Press.

Huron, D., & Hippel, P. v. (2000). 'Why do skips precede reversals? The effect of tessitura on melodic structure.' *Music Perception,* 18(1).

Kaneko, H., Kawakami, D., & Sagayama, S. (2010). 'Functional Harmony Annotation Database for Statistical Musical Analysis.' *Paper presented at the 2010 International Society for Music Information Retrieval Conference.* Utrecht.

Kirlin, P. B. (2009). 'Using Harmonic and Melodic Analyses to Automate the Initial Stages of Schenkerian Analysis.' *Paper presented at the 2009 International Society for Music Information Retrieval Conference.* Kobe.

Kirlin, P. B., & Utgoff, P. E. (2010). 'A Framework for Automated Schenkerian Analysis.' *Paper presented at the 2008 International Society for Music Information Retrieval Conference.* Philadelphia.

Kornstädt, A. (2001). 'The JRing System for Computer-Assisted Musicological Analysis.' *Paper presented at the 2001 International Society for Music Information Retrieval Conference.* Bloomington, Indiana.

Lartillot, O. (2005a). 'An Efficient Algorithm for Motivic Pattern Extraction Based on a Cognitive Modeling.' *Paper presented at the Open Workshop of Music Network.* Vienna.

Lartillot, O. (2005b). 'Efficient Extraction of Closed Motivic Patterns in Multi-Dimensional Symbolic Representations of Music.' *Paper presented at the 2005 International Society for Music Information Retrieval.* London.

Lartillot, O. (2005c). 'Multi-Dimensional Motivic Pattern Extraction Founded on Adaptive Redundancy Filtering.' *Journal of New Music Research,* 34(4).

Lartillot, O., & Saint-James, E. (2004). 'Automating Motivic Analysis through the Application of Perceptual Rules.' *Computing in Musicology, (*13).

Leon, K., & Hutchinson, W. (1983). 'Entropy as a Measure of Style: The Influence of Sample Length.' *Journal of Music Theory,* 27(1).

León, P. J. P. d., Rizo, D., & Iñesta, J. M. (2007). 'Towards a Human-friendly Melody Characterization by Automatically Induced Rules.' *Paper presented at the 8th International Conference on Music Information Retrieval.* Vienna.

Lerdahl, F., & Jackendoff, R. (1977). 'Toward a Formal Theory of Tonal Music.' *Journal of Music Theory,* 21(1).

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. London: The MIT Press.

Lester, J. (1992). *Compositional Theory in the Eighteenth Century.* Cambridge, Massachusetts: Harvard University Press.

Madsen, S. T., & Widmer, G. (2007a). 'A Complexity-Based Approach to Melody Track Identification in MIDI files.' *Paper presented at the International Workshop on Artificial Intelligence and Music.* Hyderabad, India.

Madsen, S. T., & Widmer, G. (2007b). 'Towards a Computational Model of Melody Identification in Polyphonic Music.' *Paper presented at the 20th International Joint Conference on Artificial Intelligence.* Hyderabad, India.

Manaris, B., Romero, J., MacHado, P., Krehbiel, D., Hirzel, T., Pharr, W., & Davis, R. B. (2005). 'Zipf's Law, Music Classification, and Aesthetics.' *Computer Music Journal,* 29(1).

Manaris, B., Roos, P., Machado, P., Krehbiel, D., Pellicoro, L., & Romero, J. (2007). 'A Corpus-Based Hybrid Approach to Music Analysis and Composition.' *Paper presented at the Twenty-Second AAAI Conference on Artificial Intelligence.* Vancouver.

Manaris, B., Vaughan, D., Wagner, C., Romero, J., & Davis, R. B. (2003). 'Evolutionary Music and the Zipf-Mandelbrot Law: Progress towards Developing Fitness Functions for Pleasant Music.' *Paper presented at the First European Workshop on Evolutionary Music and Art, Essex.* England.

Manzara, L. C., Witten, I. H., & James, M. (1992). 'On the Entropy of Music: An Experiment with Bach Chorale Melodies.' *Leonardo Music Journal,* 2(1)

Marsden, A. (2007). 'Automatic Derivation of Musical Structure: A Tool for Research on Schenkerian Analysis.' *Paper presented at the 2007 International Society for Music Information Retrieval Conference.* Vienna.

McConnell, G. (2009). 'An Introduction to Species Counterpoint.' Retrieved December 2nd, 2011, from http://homepage.eircom.net/~gerfmcc/SpeciesOne.html

Meehan, J. R. (1980). 'An Artificial Intelligence Approach to Tonal Music Theory.' *Computer Music Journal,* 4(2).

Meredith, D., Wiggins, G., & Lemström, K. (2002). 'Algorithms for Discovering Repeated Patterns in Multidimensional Representations of Polyphonic Music.' *Journal of New Music Research,* 31(4).

Meyer, L. B. (1957). 'Meaning in Music and Information Theory.' *The Journal of Aesthetics* *and Art Criticism,* 15(4).

Narmour, E. (1989). 'The "genetic code" of melody: Cognitive structures generated by the implication-realization model.' *Contemporary music review,* 4(1).

Narmour, E. (1990). *The Analysis and Cognition of Basis Melodic Structures*. Chicago: The University of Chicago Press.

Nelson, O. (1966). *Patterns for Improvisation*. New Albany, Indiana: Jamey Aebersold Jazz.

Nettl, B. (1958). 'Some Linguistic Approaches to Musical Analysis.' *Journal of the International Folk Music Council,* (10).

Niitsuma, M., Fujinami, T., & Tomita, Y. (2009). 'The Intersection of Computational Analysis and Music Manuscripts: A New Model for Bach Source Studies of the 21st Century.' *Paper presented at the 2009 International Society for Music Information Retrieval Conference.* Kobe.

Pachet, F. (1997). 'Computer Analysis of Jazz Chord Sequences: Is *Solar* a Blues?' In E. R. Miranda (Ed.), *Readings in Music and Artificial Intelligence*. Amsterdam: Harwood Academic Publishers.

Paiva, R. P. (2005). 'An algorithm for melody detection in polyphonic recordings.' *Paper presented at the The 1st Music Information Retrieval Evaluation eXchange.* London.

Pardo, B., & Birmingham, W. P. (2000). 'Automated Partitioning of Tonal Music.' *Paper presented at the 2000 Florida AI Research Society Conference.* Florida.

Pardo, B., & Birmingham, W. P. (2002). 'Algorithms for Chordal Analysis.' *Computer Music Journal,* 26(2).

Pearce, M. T. (2005). 'The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition.' Doctoral Thesis from City University, London.

Pearce, M. T., Meredith, D., & Wiggins, G. A. (2002). 'Motivations and Methodologies for Automation of the Compositional Process.' *Musicae Scientiae,* 6(2).

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2008). 'A Comparison of Statistical and Rule-Based Models of Melodic Segmentation.' *Paper presented at the The 9th International Society for Music Information Retrieval Conference.* Philadelphia.

Pearce, M. T., & Wiggins, G. A. (2004a). 'Improved Methods for Statistical Modelling of Monophonic Music.' *Journal of New Music Research,* 33(4).

Pearce, M. T., & Wiggins, G. A. (2004b). 'Rethinking Gestalt Influences on Melodic Expectancy.' *Paper presented at the The 8th International Conference on Music Perception & Cognition.* Evanston, Illinois.

Pearce, M. T., & Wiggins, G. A. (2006). 'The information dynamics of melodic boundary detection.' *Paper presented at the The 9th International Conference on Music Perception and Cognition.* Bologna, Italy.

Pearce, M. T., & Wiggins, G. A. (2007). 'Evaluating Cognitive Models of Musical Composition.' *Paper presented at the The 4th International Joint Workshop on Computational Creativity.* London.

Phon-Amnuaisuk, S., Smaill, A., & Wiggins, G. (2006). 'Chorale Harmonization: A View from a Search Control Perspective.' *Journal of New Music Research,* 35(4).

Powers, D. M. W. (1998). 'Applications and Explanations of Zipf's Law.' *Paper presented at the New Methods in Language Processing and Computational Natural Language Learning '98.* Sydney.

Rahn, J. (1980). 'On Some Computational Models of Music Theory.' *Computer Music Journal,* 4(2).

Raphael, C. (2003). 'Harmonic Analysis with Probabilistic Graphical Models.' *Paper presented at the 2003 International Society for Music Information Retrieval Conference.* Baltimore, Maryland.

Roads, C. (1980). 'Artificial Intelligence and Music.' *Computer Music Journal,* 4(2).

Schellenberg, E. G. (1996). 'Expectancy in Melody: Tests of the implication-realization model.' *Cognition,* 58.

Schoenberg, A. (1967). *Fundamentals of Musical Composition*. London: Faber and Faber Limited.

Smoliar, S. W. (1980). 'A Computer Aid for Schenkerian Analysis.' *Computer Music Journal,* 4(2).

Soltau, H., Schultz, T., Westphal, M., & Waibel, A. (1998). 'Recognition of music types.' *Paper presented at the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing.* Seattle.

Temperley, D. (2007). *Music and Probability.* Cambridge, Massachusetts: MIT Press.

Thom, B., Spevak, C., & Höthker, K. (2002). 'Melodic segmentation: evaluating the performance of algorithms and musical experts.' *Paper presented at the The International Computer Music Conference.* Göteborg, Sweden.

Thompson, W. F., Balkwill, L.-L., & Vernescu, R. (2000). 'Expectancies generated by recent exposure to melodic sequences.' *Memory & Cognition,* 28(4).

Thompson, W. F., Cuddy, L. L., & Plaus, C. (1997). 'Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task.' *Perception & Psychophysics,* 59(7).

Thompson, W. F., & Stainton, M. (1998). 'Expectancy in Bohemian Folk Song Melodies: Evaluation of Implicative Principles for Implicative and Closural Intervals.' *Music Perception,* 15(3).

Toiviainen, P., & Eerola, T. (2001). 'A Method for Comparitive Analysis of Folk Music Based on Musical Feature Extraction and Neural Networks.' *Paper presented at the 7th International Symposium on Systematic and Comparitive Musicology.* Jyväskylä, Finland.

Vos, P. G. (1989). 'Ascending and Descending Melodic Intervals: Statistical Findings and Their Perceptual Relevance.' *Music Perception,* 6(4).

Whorley, R., Wiggins, G. A., Rhodes, C., & Pearce, M. T. (2010). 'Development of Techniques for the Computational Modelling of Harmony.' *Paper presented at the First International Conference of Computational Creativity.* Lisbon, Portugal.

Wiggins, G. A., Müllensiefen, D., & Pearce, M. T. (2010). 'On the non-existence of music: Why music theory is a figment of the imagination.' *Musicae Scientiae,* 5.

Winograd, T. (1968). 'Linguistics and the Computer Analysis of Tonal Harmony.' *Journal of Music Theory,* 12(1).

Youngblood, J. E. (1958). 'Style as Information.' *Journal of Music Theory,* 2(1).

Appendix: All Contentious Sections in Melodic Reductions

For most of the material which was used in this research, numeric reduction is relatively straight-forward, following the guidelines established in sections 2.1 and 2.2. There are, however, several exceptions where difficulties arise. All the excerpts of the expositions used which present certain difficulties in melodic reduction are shown here, along with details of how each case was treated. While there are also numerous subjective decisions regarding the treatment of chromaticism in these melodies, it is extremely rare that chromatic notation is employed in this study (evidenced by the extremely small amount of chromatic motives present in the incidence tables). The justification for this, wherever it occurs, is to maximise the amount of material that is related purely by generalised shape. Each of the following examples, by contrast, requires specialised explanation.

Example 1: Clementi Sonata Op. 1 No. 2, bars 38-39

In this excerpt, although the arpeggio melodic figure continues into the bass, the treble line is taken as the melody. This is because it is necessary to create some (potentially arbitrary) division point between bass and treble voices, and the fact that the arpeggiation in the bass continues into a bass-line figure in bar 39 implies that the left hand is functioning as a separate voice. The numeric reduction of this excerpt is 37.5, 39.5, 37.5, 34.5, 32.5, 27.5, 30.5.

Example 2: Clementi Sonata Op. 1 No. 4, bars 19-21

Perceptually, it could be argued that the downbeats in the left hand starting on beats three and four in bar 19 could function melodically, especially since they don't sustain. However, because this figure continues underneath the melody on the third beat of bar 20, it is taken in the reduction to be a separate line which is not melodic. The numeric reduction of this excerpt is 36.5, 35.5, 44.5, 42.5, 39.5, 42.5, 39.5, 37.5, 39.5, 37.5, 35.5, 37.5, 35.5, 32.5, 30.5, 31.5, 32.5, 31.5, 33.5, 32.5, 31.5, 30.5, 31.5, 38.5, 36.5, 32.5, 33.5.

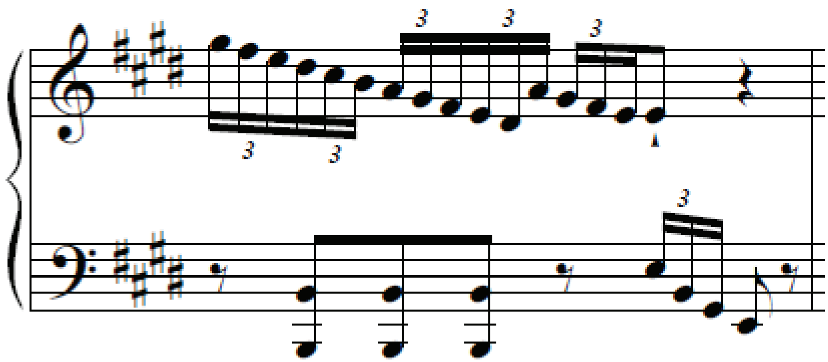
Example 3: Clementi Op. 8, No. 3, bars 34-36

There is some ambiguity here about which voices are functioning melodically. Since the descending crotchets in bar 35 step down to a sustained voice in bar 36 which is not melodic, they are taken to be part of the bass figuration, and are not included in the reduction. The octave C3 and C4 in bar 34 is treated as a continuation of bass pattern, despite appearing in a melodic 'space'. The numeric reduction of this excerpt is 43.5, 42.5, 41.5, 40.5, 38.5, 39.5, 33.5, 34.5, 37.5.

Example 4: Haydn Sonata No. 37, bar 17

In this passage, although there is an upper and lower part played between both hands, neither is sustained. This perceptually creates a single monodic structure. As such, all notes here are treated as part of a single melody. The numeric reduction of this excerpt is 38.5, 40.5, 35.5, 39.5, 36.5, 38.5, 30.5, 35.5, 31.5, 36.5, 33.5, 38.5, 32.5, 35.5, 31.5, 34.5.

Example 5: Haydn Sonata No. 46, bar 8



Although the semi-quaver triplet in the left hand at the end of this bar is a continuation of the semi-quaver triplets heard previously, it overlaps with the final note of the upper voice, and is therefore treated as a separate bass voice. The reduction of this passage is 40.5, 39.5, 38.5, 37.5, 36.5, 35.5, 34.5, 33.5, 32.5, 31.5, 30.5, 34.5, 33.5, 32.5, 31.5, 31.5.

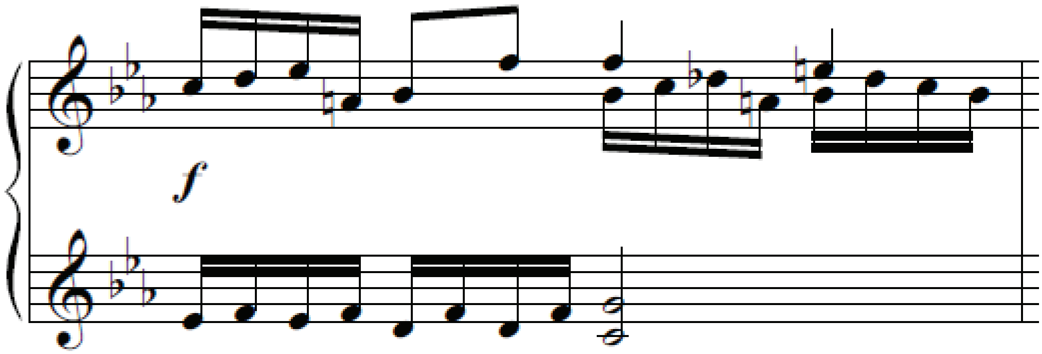
Example 6: Haydn Sonata No. 46, bars 12-13

This example shows a continuation of a line similar to that in example 1. Here, the semiquaver triplets at the end of bar 12 in the left hand are taken to be separate to the melody, because even though they continue melodic motives, they descend into a separate sustained voice. Acknowledgement must be made that this passage is open to interpretation, however. The numeric reduction of this excerpt is 35.5, 37.5, 36.5, 35.5, 37.5, 39.5, 42.5, 41.5, 40.5, 39.5, 38.5, 37.5, 37.5, 36.5, 35.5, 34.5, 33.5, 32.5, 31.5.

Example 7: Haydn Sonata No. 50, bars 30-34

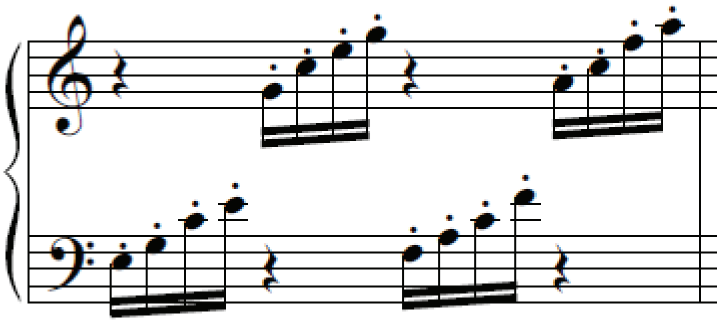
This passage shows several examples of melodic voice splitting. The semi-quavers taken over by the left hand in beat three of bar 30 overlap, and create overlap, with the semi-quavers in the right hand. They are thus taken to be part of bass figuration, and not a melodic figure. However, the change of register of the quaver figure between bars 33 and 34 does not create any overlap. In this case, all the quavers are included melodically. The numeric reduction of this excerpt is 46.5, 44.5, 46.5, 42.5, 44.5, 39.5, 42.5, 37.5, 37.5, 39.5, 42.5, 44.5, 39.5, 42.5, 44.5, 46.5, 39.5, 42.5, 44.6, 46.5, 43.5, 41.5, 43.5, 39.5, 41.5, 37.5, 39.5, 36.5, 37.5, 34.5, 36.5, 32.5, 34.5, 30.5, 32.5, 31.5, 34.5, 36.5, 38.5, 41.5, 43.5, 45.5, 24.5, 26.5, 28.5, 30.5, 42.5.

Example 8: Haydn Sonata No. 51, bar 16



It must be conceded that in this example, polyphonic imitation clearly occurs between the two upper voices. While the semiquavers in beats three and four are clearly a continuation of the preceding semiquaver melodic material, the top voice is taken to be the most perceptually significant. Apart from simply being the top sounding voice, it is is also chosen to maintain continuity and uniformity throughout the rest of the reduction. The numeric reduction of this excerpt is 36.5, 37.5, 38.5, 34.5, 35.5, 39.5, 39.5, 38.5.

Example 9: Mozart Sonata No. 1, bar 9

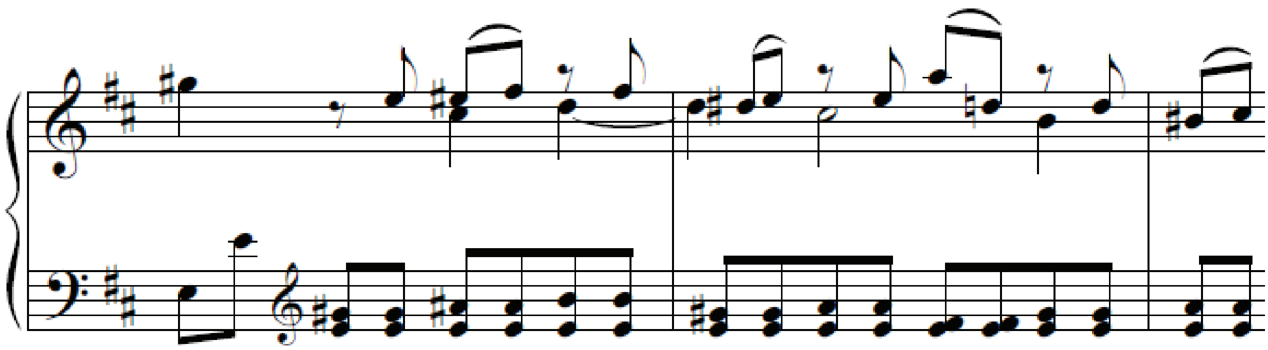


The writing here is split evenly between left and right hands, but perceptually, there is one continuous melody. In figures such as these, all notes are taken into account as part of the melody. The numeric reduction of this excerpt is 24.5, 26.5, 29.5, 31.5, 33.5, 36.5, 38.5, 40.5, 25.5, 27.5, 29.5, 32.5, 34.5, 36.5, 39.5, 41.5.

Example 10: Mozart Sonata No. 1, bars 27-28

This passage is problematic, because the quaver accompaniment figure is still within the range of the upper line, and could easily be an example of compound melody. However, this interpretation does not take into account the rhythmic overlap between the crotchets in the right hand, and the quavers in the left hand. For this reason, only the right hand is included in the melodic reduction. The numeric reduction of this excerpt is 35.5, 37.5, 36.5, 35.5, 33.5, 34.5, 35.5, 36.5, 37.5, 38.5.

Example 11: Mozart Sonata No. 6, bars 34-36



This is another excerpt, like example 8, where there are clearly two independent upper voices. The upper voice achieves prominences through its greater disjunct motion and activity, and its notes alone are taken solely as the melodic reduction. The numeric reduction of this excerpt is 40.5, 38.5, 38.75, 39.5, 39.5, 37.5, 38.5, 38.5, 41.5, 37.5, 37.5, 35.5, 36.5.

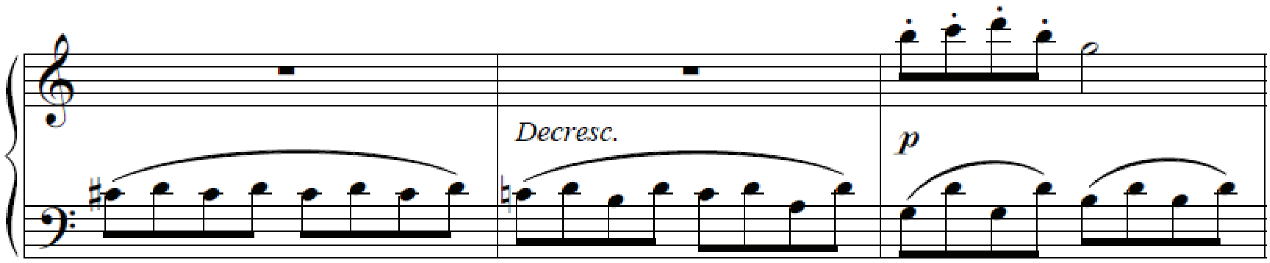
Example 12: Mozart Sonata No. 6, bar 45

In this bar, the fact that this figuration has no accompaniment precedent (unlike in example 3) means that it can function as a single disjunct melodic line. The top voice of each chord is taken to be melodic. The numeric reduction of this excerpt is 26.5, 44.5, 27.5, 43.5.

Example 13: Mozart Sonata No. 7, bars 21-22

This excerpt is included not because of any difficulty in reduction to a single melodic line, but because it serves as an example of the difficulty in numerically reducing chromatic passages. There is clearly a pattern here where chord tones are embellished with chromatic passing notes, which are all a semi-tone beneath each note of the C major triad. This figure later occurs again, on a different chord, so the chromaticism is clearly motivic. There is difficulty in representing this however — most chord tones can have their chromatic neighbour represented in numbers, but the same interval when it occurs from B to C has no unique representation. In the interest of maintaining a generic notation which can apply to all keys, the chromaticism of this passage must sadly be abandoned, turning a chromatically embellished arpeggio into a scalic motive. The numeric reduction of this excerpt is therefore 36.5, 37.5, 36.5, 35.5, 36.5, 37.5, 38.5, 39.5, 40.5, 42.5, 43.5, 44.5, 45.5.

Example 14: Mozart Sonata No. 7, bars 33-35

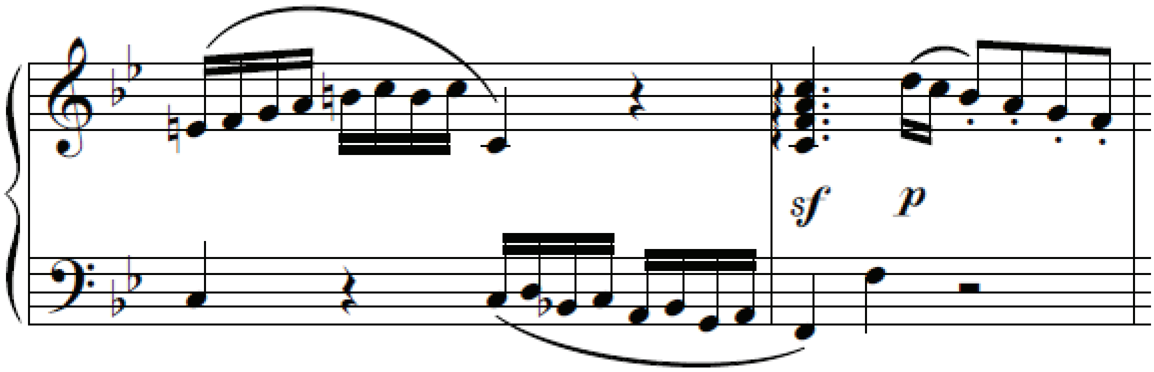


The passage in bars 33 and 34 in the bass could be taken to be melodic — it is the only voice, and it is not chordal. However, considering the continued left hand figuration of this part of the exposition, it is more likely that this passage is simply an accompaniment figure which bridges two melodies. This makes the staccato entry in the right hand in bar 35 the 'true' melody. This is the interpretation used in the melodic reduction in this thesis. As such, the numeric reduction of this excerpt is 42.5, 43.5, 44.5, 42.5, 40.5.

Example 15: Mozart Sonata No. 13, bar 22

In passages such as this, it is debatable whether the left hand octave on the third beat constitutes a continuation of the melodic arpeggiation, or a continuation of the accompaniment figure suggested by the octave in the left hand on the first beat. Since the accompaniment figure in this bar is without precedent in this exposition up to this point, and because it perceptually sounds as a continuation of the arpeggio, it is taken to be melodic. The numeric reduction of this excerpt is 38.5, 43.5, 40.5, 38.5, 36.5, 33.5, 31.5, 29.5, 22.5.

Example 16: Mozart Sonata No. 13, bars 30-31



The left hand passage in beats three and four of bar 30 presents a similar situation to that shown in example 3 and example 6. Even though the figure takes the melodic rhythm from beats one and two of bar 30, it causes rhythmic overlap with the top voice in beat three, and leads into a clearly accompanying octave on the first two beats of bar 31. For this reason, it is taken as accompaniment figuration and is not included in the reduction. The numeric reduction of this excerpt is 31.5, 32.5, 33.5, 34.5, 35.5, 36.5, 35.5, 36.5, 29.5, 36.5, 37.5, 36.5, 35.5, 34.5, 33.5, 32.5.

Example 17: Mozart Sonata No. 13, bars 43-44 and 46-47

The two passages above are clearly rhythmically related. Identifying a single melody in them is not straightforward — in the first excerpt, the top voice could be a accompanying figure for the rising pattern in the left hand. It is equally valid to say that the quavers form an accompaniment for the melody in minums above them. Identification becomes even trickier in bars 46-47, where the voicing is reversed. In this situation, the somewhat arbitrary decision has been made to isolate the top voice as melodic, citing its perceptual significance. Therefore, the numeric reduction of bars 43-44 is 44.5, 38.5, 42.5, 41.5, and the numeric reduction of bars 46-47 is 39.5, 38.5, 39.5, 38.5, 37.5, 38.5, 37.5, 44.5, 44.5, 44.5, 37.5, 37.5.