Prosody beyond pitch and emotion in speech and music: evidence from right hemisphere brain damage and congenital amusia

Ariadni Despoina Loutrari

December 2015

52,839 words

Submitted in partial fulfilment of the Requirements for the Degree of:

Doctor of Philosophy in Applied Linguistics

Birkbeck, University of London

Declaration

I declare that the work presented in this thesis is my own.

Ariadni Despoina Loutrari

Declaration

Abstract

This dissertation examines the relationship of prosodic processing in language and music from a new perspective, considering acoustic features that have not been studied before in the framework of the parallel study of language and music. These features are argued to contribute to the effect of 'expressiveness' which is here defined as the combination of the acoustic features (variation in duration, pitch, loudness, and articulation) that results in aesthetic appreciation of the linguistic and the musical acoustic stream and which is distinct from pitch, emotional and pragmatic prosody as well as syntactic structure.

The present investigation took a neuropsychological approach, comparing the performance of a right temporo-parietal stroke patient IB; a congenitally amusic individual, BZ; and 24 control participants with and without musical training. Apart from the main focus on the perception of 'expressiveness', additional aspects of language and music perception were studied. A new battery was designed that consisted of 8 tasks; 'speech prosody detection', 'expressive speech prosody', 'expressive music prosody', 'emotional speech prosody', 'emotional music prosody, 'speech pitch', 'speech rate', and 'music tempo'. These tasks addressed both theoretical and methodological issues in this comparative cognitive framework.

IB's performance on the expressive speech prosody task revealed a severe perceptual impairment, whereas his performance on the analogous music task examining 'expressiveness' was unimpaired. BZ also performed successfully on the same music task despite being characterised as congenital amusic by an earlier study. Musically untrained controls also had a successful performance. The data from IB suggest that speech and music stimuli encompassing similar features are not necessarily processed by the same mechanisms. These results can have further implications for the approach to the relationship of language and music within the study of cognitive deficits.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am deeply grateful to my supervisor, Professor Marjorie Lorch, for her valuable guidance and encouragement throughout all stages of my doctoral dissertation. Marjorie's scholarship and support have been fundamental to the formulation of my research questions, the collection of data, the subsequent analysis, and the writing process. I feel very fortunate to have had such a mentor along this long journey and I consider myself hugely privileged to be her student.

I am very thankful to Harry Elektron for his help in creating the auditory stimuli and his patience during the design of my research tools. I would also like to thank Evangelos Paraskevopoulos who provided me the opportunity to meet and test BZ. I am indebted to Dr Harriet Proios for her support and feedback during the data collection in Renaissance Rehabilitation Centre, in Greece, as well as the speech therapists of her Department. I am also particularly thankful to Professor Lazaros Triarhou whose kind help before the data collection process was critical to the completion of my study.

I would also like to thank all the participants of the study for their precious time and help. I am particularly grateful to IB for his time and patience throughout the testing sessions.

1. Introduction

This chapter introduces the background of the present investigation looking at literature in the field of the parallel study of language and music from a neuropsychological perspective. Starting with more general questions that have been addressed in the field, the reader is introduced to the questions that are specific to this study. After a short introduction on the historical roots of the language-music relationship, references are made to comparisons between those variables that appear to have received the most attention in previous investigations. As a great deal of research has focused on the role of pitch in the processing of language and music and their possible connections through pitch, this chapter refers to studies that addressed relevant questions and then turns to issues pertaining to theories of cognition and how these apply to this field. Moreover, it examines the evidence from language and music impairments and what these can reveal for the investigation of these two cognitive faculties when one compares the 'normal' brain with the lesioned brain. References are made to the different types of associations and dissociations in the performance of brain-damaged individuals that have been reported by researchers as well as to some relevant interpretations.

In order to appreciate the contribution of this study, it is worthwhile to take into account broader issues that have been raised in the field. Researchers have addressed a large number of questions. This literature review will look at some crucial questions in the framework of the language-music relationship. Is the music-language relationship investigation a trend of only contemporary research? How

does culture affect the development of linguistic and musical abilities? Are the skills that one acquires through music training transferable to language and vice versa? Does intensive training result in structural and functional plasticity in the brain? Are there any aspects of speech and music that bear more similarities compared to others? If pitch contours constitute one of these features that can be considered analogous in speech and music, are there further aspects of pitch differing across domains in more narrowed-down investigations? Which theories of cognition are relevant to the parallel study of language and music? What can neurological disorders reveal about the processing of speech and music? As this study investigated perception rather than production abilities, the review of previous studies also focuses on how one perceives speech and music input rather than on how they produce analogous output.

It is worth noting that the terms 'language' and 'speech' are not used interchangeably here. In the following chapters, the term 'speech' is used when referring to perceptual processes or acoustic properties of spoken language. Speech can be defined as a sound stream organised in several processing stages that relate to stored representations which are employed while the listener is exposed to a linguistic stimulus (Scott and Johnsrude, 2003). In more generic references and comparisons, the term 'language' is used to indicate all of the cognitive components of the linguistic domain in distinguishing it with the music domain more generally. With respect to music, all references in the present dissertation have to do with Western tonal music whose conventions do not necessarily apply to all known musical systems.

It should be also noted that the literature review of this dissertation does not follow some of the established conventions of critical analysis. There are no comparisons

among studies discussed here, except for a few cases. The interdisciplinary nature of this work does not allow for direct comparisons among the reviewed investigations. This is due to several reasons. As this dissertation employs disparate sources of evidence with different frameworks, different participants and distinct research objectives, the conclusions of all these studies cannot be compared with one another. That is, as the research questions discussed draw on several fields of research that often focus only on one of the components of the present investigation, one-to-one comparisons are rendered impossible. The purpose of the literature review is to rather put these disparate strands of evidence together in order to bridge the existing gap. This can be better understood given the fact that the questions addressed here did not derive from direct consideration of the existing comparisons in the literature and do not correspond to equivalent entities in previous language-music comparisons.

1.1 The study of language and music: before contemporary investigations

The study of language and music in a comparative framework has been particularly fruitful during the last decades. However, comparisons between the linguistic and the musical domain had been drawn much earlier through focused specific comparisons and by more indirect observations. Moreover, the terminology that has been used historically for some concepts is also indicative of the relationship that was detected between language and music before systematic study of the two cognitive domains began.

The etymology of speech prosody terms indicates its connection with music. It provides interesting indirect connections that reveal consideration of speech 'in

musical terms'. Besides denoting a way of pronunciation, the word 'accent' indicates acoustic stress and emphasis. In Latin, 'accentus' meant song in speech, as occurred in the musical dimension of verse, as the word 'cantus' signifies singing (Oxford English Dictionary, 2015). Delving deeper into the evolutionary journey of the word, as it is shown in the same dictionary entry, the word 'accent' is a loan-translation of the word 'prosoidia' which originally meant pitch variation in the acoustic realisations of written poetry. Although the practice of poetry is not representative of a systematic investigation between linguistic and musical relationships, it reveals intuitions about this relationship.

The acoustic organisation of classical Greek and Latin shows the effect that poetry had on ancient cultures. For example, rhythmic relationships in the Greek and the Latin language seem to have emerged out of the appreciation of poetry and theatre (Bolton, 1894). That is, although these patterns were not naturally occurring in the language, musical aspects of poetry gave speech this additional dimension. Anderson (1973) refers to an example of a theoretical consideration of the language-music relationship in the ancient world. A 4th century BCE pupil of Aristotle, Aristoxenus, observed that speaking and singing display two different realisations of pitch. In spoken Greek, he characterised the transition between pitch events as continuous (syneches). In contrast, according to Aristoxenus, pitch movements in singing differ in that the voice pauses on them (diastematike), resulting in a different acoustic effect.

Apart from descriptions of acoustic realisations of speech and music forms, issues on the origin of language and music remain debatable in present days. However, the debates on the origin of language and music have their own origin. From the 17th

century onwards, many debates on the human origin of language and music took place (Besson and Shon, 2003).

The study of patients with brain damage and the use of neuroimaging techniques in order to identify common neural substrates between language and music processes is currently providing large amounts of data. In the 19th century, neurologists displayed a similar interest in understanding how language and music are represented in the brain and what links between them meant. In fact, studying language and music was not a research objective *per se*, as often the interest of 19th century neurologists' in understanding language and music derived from the interest in discovering the complexities of human cognition (Johnson et al., 2010). Hence, the study of the compromised cognitive system occurred in the general framework of the study of brain and human behaviour. As the authors note, neurologists used to construct diagrammatic models after observing and analysing the behaviour of their patients. This happened in order for them to show schematically their observations about language processing. Similar to those diagrams about language, some years later in the same century, music processing diagrams were also designed. This piece of evidence reveals a link between clinical practice and research which is of paramount importance in present days.

Several speech disorders are now grouped under the umbrella term 'aphasia'. This term was initially coined as 'aphemie' (aphemia) by Paul Broca whose contribution seems to be overrated in contrast to other clinicians that are not currently sufficiently acknowledged in the literature (Lorch, 2011). Broca's name is associated with *Broca's aphasia*, inability to produce speech. In present days, the patterns of spared and compromised linguistic behavior are shown to be much more complex, as it is known that there is not an absolute dichotomy between production and perception

deficits. For example, patients with Broca's aphasia display some comprehension difficulties as well. However, the contribution of 19th century research was critical to the study of linguistic disorders. It was during the same century that music disorders attracted the interest of investigators. The French neurologist Jean-Martin Charcot examined the musical abilities of his patients and discussed issues on music localization in the brain as well as the relationship between language and music impairments (Johnson et al., 2013). It, therefore, seems that although the study of language had received earlier attention than music, the two mental faculties had received a great deal of attention during the 19th century and they do not constitute a modern research trend.

Johnson et al. (2010) discuss the work of prominent neurologists such as John Hughlings Jackson and Jean-Baptiste Bouillaud with individuals that displayed interesting patterns of linguistic and musical abilities. Such cases include, for example, the descriptions of patients with aphasia that had intact singing ability. This observation by John Hughlings Jackson was followed by the conclusion that the vocal articulation and control of muscles is intact in aphasic patients, as their singing abilities showed. Given that these are indispensable for singing, it was argued that something else must have impeded the production of speech in such patients. As it will be discussed in 1.4, modularity has been an influential concept in 20th century cognitive science. This concept seems to have been captured in some way many years before it was proposed as a theory by Fodor (1983). In an attempt to explain the signing abilities of a patient in the presence of impaired speaking abilities, John Hughlings Jackson proposed that there are two branches of linguistic ability; one of them was named 'emotional language' and the other 'intellectual language'. Emotional language was often spared in individuals with severely impaired speech

production and in some cases the ability to produce song was also preserved (Lorch and Greenblatt, 2015). By contrast, intellectual language was not spared in aphasic patients. More interestingly, Henry Charlton Bastian provided detailed descriptions on the performance of patients on several modalities on linguistic perception and production including music and insisted on the systematic assessment of aphasic patients (Lorch, 2013). Such 19th century investigations into the spared and impaired abilities of aphasic patients made a major contribution to the understanding of the modular organisation of language function, and its components which are not necessarily compromised at the same time in the presence of brain damage.

The design of a therapy tool called 'Melodic Intonation Therapy' (Albert et al. 1973) has been based on observations made one century earlier. This tool capitalises on the preserved singing abilities of aphasic patients, something that was first discovered in the 19th century.

Modern research has focused on the relationship between language and music in a more systematic way. Researchers from several different fields such as neurolinguistics, cognitive science, and neuropsychology have been participating in a very fruitful interdisciplinary dialogue for the past decades. It is, however, important to take into account the fact that current research has built upon previous knowledge which was obtained throughout several centuries.

1.2 General comparisons

From general to narrow comparisons, language and music display various similarities and differences. This section refers to some comparisons, mainly from a cognitive standpoint which is more closely related to the neuropsychological approach of this study. Although the biological framework in which the two faculties

developed has been less clearly delineated, the recent research in the study of transferable skills from one domain to the other and the study of perception of linguistic and musical acoustic features have shed some light on the relationship between the two faculties.

1.2.1 The language-music relationship from a cultural, evolutionary, and cognitive perspective

From a cultural perspective, both language and music develop within a certain context. That is, humans' cognitive processing abilities are shaped by the linguistic and the musical culture to which they belong and they both participate in the creation of cultural traditions including poetry and songs (Besson and Shon, 2003). More interestingly, language and music have been argued to shape each other to some degree due to the cultural context that they share. Patel et al. (2006) studied the music of English and French composers belonging to musical nationalism era (1800s-1900s) and found that their music reflected durational contrastiveness patterns of their native language. In other words, their results suggested that the music of the composers reflected the prosody of their native language. Jackendoff (2009) argues that using language and music is interwoven with social and cultural interaction and that all cultures possess both a local variant of language and a local variant of music, thus building culturally shared repertoires in both domains of a given social group. It is argued that, being exposed to specific pitch patterns of a musical culture, listeners tend to internalise these patterns which, at the same time, affect their music experience as they form expectations on which notes are likely to follow a pitch sequence before they are actually played (McDermott and Oxenham, 2008). Variation due to cultural factors is present both at the linguistic and the

musical level and the two faculties cannot be dissociated completely from this cultural context.

Both language and music can be examined in an evolutionary framework that seems to provide a general link between the domains. Peretz (2006) argues that there is a good reason to assume that music is based on principles guided by innate mechanisms, in contrast to invention of codes (e.g. the Morse code) transmitted by nongenetic mechanisms. Many researchers hold the view that language evolution was based on some pre-adaptations in the human lineage that related to joint attention, meaningful use of arbitrary symbols, representation of complicated concepts, and improved memory for sound sequences (Christiansen and Kirby, 2003). In music, mechanisms shared by animals suggest that musical processes were also based on foundational resources common to both humans and other primates (Fitch, 2006). There is evidence that the auditory system of a given species adapts to the sound processing requirements of their environment (e.g., bat ecolocation). Zatorre et al. (2002) suggest that humans can be deemed as a species with a nervous system adapted to their linguistic and musical environment. Research on the biological foundation of linguistic and musical abilities can, thus, provide answers to more general biological questions both with respect to the evolution of cognitive abilities as well as their possible relationships.

An examination of language-music similarities and differences from a cognitive perspective provides more direct means of investigating their relation in human auditory processing. Nevertheless, a cognitive study of speech and musical relationships is not without hurdles. Using a neuropsychological perspective, speech processing can be examined through comparing the impaired performance in individuals with neurological compromise with the expected 'normal' performance

of healthy individuals due to a large research base on speech perception processes which has built up for over 50 years. In contrast, investigation of music processing disorders is handicapped by a more limited pool of evidence regarding the neurotypical musical brain (Stewart et al., 2006). The cognitive psychology of music is a relatively young field of empirical research which renders comparisons between the two domains a rather challenging task.

1.2.2 Perceptual features of language and music

When it comes to the perceptual features of language and music, one can easily tell that some elements bring the two domains much closer in comparison to other features. There are some elements that do not appear to have counterparts across domains. For instance, a clearly differentiating feature between language and music is semanticity. While words can be analysed in terms of semantic features that are defined by arbitrary connections between sounds and meaning, music has no extramusical space that can assign meaning to notes or chords (Besson and Shon, 2003). It appears that music has meaning in the sense that it can arouse and inhibit emotional response (Meyer, 1956). In this sense, it can be argued that they both employ modulations of an acoustic signal to convey some type of information (Zatorre et al., 2002). However, the information conveyed by music is relatively limited and differs from the semantic complexity that is present in language and the sophisticated communication function that it serves.

On a general level, both language and music are rule-based systems that are composed of basic structural elements that are synthesised to form higher-order structures, according to syntax and harmony respectively (Besson and Shon, 2003). As language has phonemes that are combined into utterances and form part of a

larger discourse, so too, music has notes that are combined into music phrases which form part of longer musical passages. In order to comprehend these complex acoustic sequences, the structural elements of each domain (words and notes or chords) appear to be governed by rules that determine their relationship with preceding events (Patel, 2012).

Both language and music include sequences of patterns in their repertoire (Jackendoff, 2009). Moreover, pitch variations and rhythm are present in both the speech and the music signal. However, there are notable differences across domains. Intonation contours [pitch contours extending over chunks of utterances, (Meyer et al., 2004)] appear to constitute a language component that can be more closely related to music, despite the fact that it does not consist of specific tone sets found in music and it does not comply with music scales rules (Patel et al., 2005). Put a different way, despite the similarity in the organisation of the acoustic stream of language and music in pitch contours at the macro level, human auditory processes for detecting acoustic patterns in speech and music display differences. Music melodies are created by concrete pitch patterns, while speech contours are only perceived as a relative rise and fall of the pitch (Jackendoff, 2009). In other words, one perceives pitch directions across an utterance and not specific values between pitch events (Johhsrude et al., 2000).

In order to better appreciate this difference between language and music, it is worth taking a closer look at the concept of musical key. In Western music, musical events develop in a specific key context. A good metaphor to explain this is that there is always a centre of gravity and some chords are more stable and others are less stable in a given key context (Tillmann and Bharucha, 2000). A key can be also defined as a probability distribution that allows certain notes within a pitch set to occur more

frequently than others and to have a longer duration (Mc Dermott and Oxenham, 2008). It seems that there is not an analogous contour organisation in language. Speech contours do not seem to be based on similar rules that only allow for specific intervals and there are no rules imposing the occurrence of a specific pitch in contrast to others. In music, the existence of more complex pitch relationships might demand the use of more refined pitch mechanisms in contrast to speech (Zatorre and Baum, 2012). This difference in pitch complexity between these cognitive faculties suggests that the human ear might not be as sensitive to small manipulations of speech contours, whereas small manipulations in musical pitch can be perceived as not belonging to the right key context. This ability to detect mistakes in key has been shown for both musicians and non-musicians (Tillmann and Bharucha, 2000, Warrier and Zatorre, 2002). Hence, it appears that similar general principles exist across domains but these principles are expressed in different patterns or, put it a different way, in different degrees of accuracy.

Pitch constitutes a major building block of recursion in music, whereas language does not depend on pitch in order to construct computational hierarchies. In the domain of language, one can make use of a limited set of elements in order to create an infinite number of expressions that convey semantic meaning (Hauser et al., 2002). Smaller structures are embedded in larger structures having a direct effect on the meaning of a given sentence. The way these computational procedures function through the use of syntactic relationships are argued to distinguish human language from animal communication systems (Corballis, 2007). A conjunction, for example, can introduce an additional phrase embedded in a previous phrase and this pattern can be repeated, often resulting in a very complex output. In Western tonal music, recursion is also argued to occur (e.g., Boltz and Jones, 1986; Jackendoff, 2009) but

syntactic relationships involving parts of speech such as subject, verb, or object do not exist. Rather, pitch seems to allow for music hierarchical structures, signalling which notes are more important than others in a given melody. Some pitches are structurally more important, whereas others have a more embellishing role, and once removed, the structurally significant pitch events are perceptually salient, not depending on those that serve an embellishing role (Lerdahl & Jackendoff, 1983). Pitch seems to be at the core of musical recursion in contrast to linguistic recursion. In general, hierarchical structure seems to depend on different elements in language and music.

A similar picture emerges in the case of rhythm. Rhythm refers to the patterns of duration and accentuation of musical phrases. Both domains display rhythmic patterns, but in musical rhythm, one can identify concrete and predictable patterns, whereas language rhythm cannot be objectively described and does not comply with analogous patterns (Schreuder, 2006). Past investigations claiming that speech displays units that occur periodically have not supported empirically this hypothesis in contrast to music studies that point to regulated periodic patterns (Patel, 2003). The recurrence of stressed events at regular intervals can be easily captured in the way one moves to the rhythm of songs, whereas something similar does not happen with speech in this systematic fashion.

Rhythmic patterns seem to be more comparable at the level of grouping. In language, words are grouped with other words in grammatical units of sequences with, strong and weak syllables, differences in speaking rate and also pauses that can signal prosodic breaks (Wightman and Ostendorf, 1994). At this more abstract level, that is, without taking the tonal structure of music into consideration, it seems that there are some clear similarities with music which can be examined without

including those features that are specific to each domain. At an even more general and abstract level, one can simply think of the fact that phrases in both language and music have the same start, middle and end structure, despite of the differences that signal the parts of these structures across domains.

1.2.3 Are music skills transferable to the speech domain?

As speech and music have certain similarities at the perceptual level, the question arises as to whether training in one domain can benefit the other. The relationship between language and music has been explored through the study of training benefits, i.e. transferable skills from one domain to the other. More specifically, a number of researchers have investigated whether music training can result in better speech perceptual skills.

Schön et al. (2004) found that training in music can have a beneficial effect to the processing of speech prosody, specifically to the judgement of pitch violations across domains. In the same line, Wong et al. (2007) showed that music experience is likely to promote processing of sounds outside the musical domain. Additionally, the ability to encode pitch can transfer from one domain to the other when it relates to acoustic features that are shared in the two domains (Bidelman and Krishnan, 2009).

Music training has been claimed to account for more successful acquisition of acoustic features in a second language. Individuals with music training appear more able to perceive stress differences than those that do not have music experience even when these acoustic patterns are absent from their native language (Kolinsky et al., 2009). In this study, the authors show that there is a positive correlation between music experience and sensitivity to stress contrasts. At the same time, however, they acknowledge the fact that music training and enhanced performance on tests do not

constitute a straightforward correlation. More specifically, they argue that taking music lessons can result in enhanced executive function. If this is the case, the result that emerges from studies comparing musicians and non-musicians does not necessarily mean that it is music experience *per se* that leads to better intonation processing abilities. Rather, a musician might be able to focus more effectively when performing a given task. It, therefore, seems that one cannot exactly determine which specific features of music knowledge –if any– result in better intonation processing skills.

In an alternative method to study the effect of music training on cognition, some researchers have looked at the plastic effects that appear to result from music lessons. In general, the brain can be modified structurally and functionally after exposure to some specific type of repetitive activity for a certain period of time. This does not necessarily occur before a specific critical period but it can also have an effect on the adult brain (e.g., Schon et al., 2004, Kraus and Chandrasekaran, 2010). There are multiple advantages for the way one responds to sound in general. For instance, White-Schwoch et al. (2013) found that the amount of music training correlated positively with the speed of neural timing in response to phoneme transitions presented in noise. An interesting point is that the training that induces plasticity does not need to be particularly long in duration in order for structural and functional similarities to be observed. Studies have shown that grey matter changes might be detectable after three months of intensive training in humans (Draganski et al. 2006). With regard to musical training specifically, Herdener et al. (2010) found that functional plasticity changes were observed in music academy students after intensive ear training that lasted for two semesters.

While these studies indicate that musical training may lead to cognitive changes under certain conditions, the present study will consider whether some aspects of music appreciation may be independent of formal training.

1.2.4 Developmental stages of language and music?

Human beings naturally develop linguistic and musical abilities except for rare cases of congenital deficits or acquired disorders later in life. Simple exposure to speech and music seems to account for a great deal of linguistic and musical cognitive abilities that are shared universally by humans. However, although learning processing seems to take place in both faculties, music learning cannot be seen as a counterpart of first language acquisition.

In both domains, human beings acquire productive skills after a stage of perception of linguistic and music features and their production reflects some patterns that they have unconsciously mastered (Besson and Shon, 2003). In other words, human infants initially engage in a receptive phase across domains before they start producing speech and music themselves. The nature of the output they produce at a later stage reflects the input they had initially received. The relationship between perception and production appears to have strong links both in speech and music through the development of motor skills. For example, by holding representations in their memory, infants manage to imitate vocal patterns successfully later in life (Kuhl, 2000). Music displays a similar characteristic that might be related to developmental processes. That is, the auditory and motor systems interact leading to a constant sensory-motor interplay (Zatorre et al., 2007). Hence, it seems that perception and production processes are inextricably interwoven when learning takes place in both language and music.

Differences, however, emerge if one takes into account the fact that music perception does not develop in clear universal stages, as it has been shown for first language acquisition. For instance, infants appear to process intonational patterns before producing two-word utterances (Snow and Balog, 2002) and also rely on acoustic cues in order to segment speech (Jusczyk and Krumhansl, 1993). Music psychology studies do not provide evidence pointing to music learning stages analogous to those of language acquisition. Moreover, opportunity of music training varies across individuals and leads to different levels of attainment in musical performance, while all healthy individuals who grow up in a speech environment will achieve fluent and grammatical language abilities. Hence, it seems that both the developmental and neuropsychological evidence are more clearly detailed in the linguistic rather than the musical domain. That is, although both faculties encompass shared elements of learning such as 'imitation', no psychological processes have been reported in music that are analogous to language learning processes such as the stages of babbling or acquisition of morphemes. In other words, learning in the musical domain does not appear to display a concrete set of component processes analogous to language.

It is not known whether musical prosody, for example, has a similar role to prosody in language in terms of statistical regularities (Palmer and Hutchins, 2006). It remains to be seen if the statistical learning that takes place in first language acquisition has a counterpart mechanism in 'music acquisition'. While language acquisition may be argued to be based on innate mechanisms that result in a similar acquisition process and a similar level of mastery for all healthy individuals, it is not clear that this is the case in the domain of music. Alternatively, if there was a 'music acquisition device', an innate mechanism for music acquisition, all listeners would

be expected to have a similar response to music stimuli and to develop music skills analogous to those of first language learners.

1.2.5 Interim Summary

Despite clear differences, language and music have some parallel properties with regard to pitch patterns. While these are present in both language and music, there is evidence that they are organised in distinct ways. The two faculties have been compared at a number of levels. Research has suggested both cultural and biological associations between language and music. Musical traditions have been argued to be affected by linguistic features. From a biological perspective, both domains are also argued to be guided by innate mechanisms and some abilities in both domains appear do develop by simple exposure. However, language acquisition has more tangible and predictable development stages of universal character, whereas not the same can be argued for music. Also, according to some studies, training in one domain may enhance processing in the other domain.

In general, there are features that appear to be closer to each other across domains and others that do not seem to have counterparts. Music units do not display any semantic dimension, that is, musical notes and phrases do not correspond, for example, to concepts, objects, or descriptions of actions as it happens in the case of speech. In terms of structure, both language and music depend on recursive patterns but realisations of these patterns seem to differ to a large degree across domains. In addition there is a difference in the status of individual components, as one cannot readily identify morphological counterparts of nouns, verbs, or adjective in the notes and chords of the musical domain. The comparisons that appear to be more productive have to do with pitch and rhythm. However, even these comparisons

should be made with caution as differences also exist. Intonation contours in speech pertain to large intervals and pitch patterns in music are much more fine-grained and governed by principles posed by scales. In the case of rhythm, both domains display rhythmic patterns, but music rhythm has a periodic character that is absent from speech. Hence, not all comparisons are equally feasible and meaningful. Culturally, biologically, and cognitively, similarities and differences have been sought in research. Challenges remain, as investigating mechanisms involved and behavioural manifestations of these mechanisms are not always easy to study.

The next section will review relevant studies in language and music in terms of pitch organisation. This will provide a foundation for the investigation of the perception of prosodic features in music and speech which is the focus of the present study.

1.3 Pitch perception across the domains of language and music

This section examines language and music through one of the most well studied features they have in common: pitch. It looks at the similarities and differences of the language and music faculties through the perceptual feature of pitch, referring later to some examples of pitch research on language and music from the angle of processing and processing disorders. It is worth reminding the reader that all the evidence presented in this section, as well as in the rest of this dissertation, concern perception and not production of pitch patterns. The reason for constraining the scope of reviewing the literature is that perception and production of identical acoustic features has been shown to relate to different processing mechanisms and brain areas. For instance, Shapiro and Danly (1985) found that the speech output of right-central and right-anterior brain-damaged patients displayed limited pitch

variation, whereas right posterior brain injured individuals manifested exaggerated pitch variation in the production of declarative, interrogative, happy or sad speech utterances. However, data on perception, as it will be discussed in 1.6 appear to involve different brain areas and processing networks.

A critical difference between the realisation of pitch patterns in speech and music was already reported in antiquity by Aristoxenus, one of Aristotle's students. As discussed in 1.1, Aristoxenus drew a distinction between the nature of the acoustic stream in speech and singing with respect to pitch. In speech, the transition from one pitch event to the next is continuous, whereas, in singing, there is a pause on all intervals (Anderson, 1973). Likewise, in recent research, it is also suggested that, in speech, pitch contours are continuous, whereas in singing the sequences of notes in a melody are discrete and 'stair-stepped' (Bidelman and Krishnan, 2009). Using musical instruments or the human voice itself, musical pitch is characterised by fixed-pitches with scale relations and only birds produce glides and other variable-frequency notes in their song (Fitch, 2006). Put a different way, in speech, the transition from one pitch to next displays some continuity, whereas in music there seems to be a 'pause' in every interval. That is, in speech, the transition from one pitch event to the next is smoother, whereas, in music, pitch events appear more discrete.

Other pitch differences seem to be related to the appreciation of linguistic and musical output. Zatorre and Baum (2012) note that pitch in music pertains to a high degree of accuracy. In contrast, experimental deviations from an original intonation contour in speech similar to manipulation applied to music contours are not perceived as unnatural by the human ear. Slight pitch manipulations in music can also have an effect on the perceived emotion, whereas a similar effect has not been

reported for speech. In melodies of the Western tonal music tradition, a small manipulation of a semitone can shift a major scale into a minor scale, thus changing the mood of the melody from happy to sad (McDermott et al., 2008).

In addition, pitch range patterns are different across domains. The distance between two pitch events is referred to as an 'interval' which is normally counted in 'tones' or 'semitones' in Western music. The smallest possible interval in the Western musical tradition is referred to as a 'semitone'. In music, intervals of 1 or 2 semitones seem to be the most frequent pitch patterns (Vos and Troost, 1989). By contrast, humans employ much larger intervals when speaking. Pitch patterns in spoken language employed to express intonation marking can go beyond 12 semitones (Fitzsimons et al., 2001). In many non-tonal languages, such as English, rising intonation at the end of an utterance signals a question whereas falling intonation is used for statements. However, in languages which employ pitch changes to distinguish different phonemic categories, so called tonal languages, such as Mandarin Chinese, there is a lot of disagreement on how intonational pitch contrasts signalling statements and questions are perceived (Yuan et al., 2002).

As discussed in 1.2.2, music pitches are organised around a 'centre of gravity' that affects their position and frequency of occurrence in a melody and compared to speech pitches, they demand more refined processing mechanisms. The specific versus non-specific organisation of pitch across domains pertains to the issue of 'tonality' which shows a critical difference between spoken utterances and melodies. Tonality can be defined as the assignment of structural significance to specific pitches in musical piece according to templates of intervallic relationships (McDermott and Oxenham, 2008). As pitch contours in speech are simply created by relative 'rises' and 'falls', it seems that pitch organisation in music is a much more

complex process. In Western tonal music, a melody is defined as a certain temporal set of pitch patterns in a musical key (Jones and Ralston, 1991). In other words, music sequences pertain to specific pitch and temporal arrangements (Steinke et al., 2001). The musical key upon which a given melody is constructed imposes strict patterns on pitch organisation in music, whereas speech intervals are not governed by such rules. At the same time, an individual will be more accurate in recognising the direction of a pitch contour rather than its specific value (McDemott and Oxenham, 2008). In order for one to perceive if an utterance corresponds to a question, they only need to perceive its relative direction of change rather than the exact value of the interval that signals this question. Timing in speech is also perceived in a more crude way compared to music. According to Fitch (2006), the fact that music depends on structured patterns of pitch and duration, whereas, in speech, pitch and time can largely vary is the most important difference between language and music organisation. Therefore, speech can be considered acoustically less predictable as compared to music.

At a more general level, the study of pitch constitutes a particularly popular research focus possibly due to the implications of its study at a broader level. The 'wetware' employed for pitch perception is likely to serve additional perceptual functions and to contribute to our knowledge about hearing in general (De Cheveigne, 2005). It should be noted that all references to pitch here have to do with relative and not absolute pitch (also referred to as 'perfect pitch') (see also 2.1.3). Relative pitch refers to the relationship between different pitches rather than the pitches themselves. A listener does not respond to absolute differences in frequency but rather to ratios between different frequencies (Lehiste, 1970). Relative pitch is critical in both music perception but also speech intonation perception, as the rise in pitch when one, for

example, forms a question is a pitch pattern perceived by the listener irrespective of the different absolute pitches across speakers (McDermott and Oxenham, 2008). Hence, references to pitch in previous sections as well as the following ones, concern the relative rather than the absolute aspect of pitch processing.

It seems that it is not pitch *per se* but rather one's sensitivity to pitch changes in an acoustic stimulus that accounts for pitch perception (De Cheveigne, 2005). In speech, differences in pitch contours contribute to the listener's ability to differentiate statements from questions or perceive different emotional tones, regardless of the semantic content of a given utterance. The patterns of rising and falling pitches of human voice and musical instruments display a significant similarity in perceptual terms. From a developmental perspective, speech prosody and music patterns are perceived early in life. In the speech domain, grammatical structures are included in suprasegmental structures which facilitate identification of phrase boundaries (Scott and Cutler, 1984). Infants are successful in segmenting speech thanks to acoustic cues that accompany syntactical structures (Jusczyk and Krumhansl, 1993). In other words, speech melody but also music processing appears to be a very early process in humans, as studies of the human fetus have demonstrated. DeCasper and Spence (1986) note that, in the third trimester of gestation, prenatal auditory experience of a foetus shaped by the mother's voice affects speech perception after birth. Similar response of third-trimester foetuses is also reported for music. Exposure of the foetus to music has revealed heart rate increase (Kisilevsky et al. 2004).

The role of statistical learning, also called experience-dependent learning, seems to provide an additional link between language and music at the level of pitch processing. Statistical learning stemming from auditory input appears to be a

powerful mechanism both in language and music acquisition. As far as language is concerned, statistical learning relates to the detection of statistical relationships between, for example, neighbouring sounds for which infants have been shown to possess effective psychological mechanisms (Saffran et al., 1996). Similarly, in music, statistical learning seems to occur even when a child is exposed to musical stimuli outside the musical system of their cultural community (Patel, 2008).

Another feature that pitch contours in both domains have in common is the fact that they are transposable; a listener will judge two melodies as the same when they start from different notes (one being higher and the other lower) and, at the same time, an utterance spoken by a male and a female will be judged as the same, despite the fact that the male utterance can be one octave lower than the female utterance (Fitch, 2006). Or, in the case of music, two melodies will be perceived as the same if the singer starts the second melody from a different note but they follow exactly the pitch patterns of the original melody. Familiar melodies can be recognised in all different transpositions, that is, when all notes of a familiar melody go up or down in pitch (McDermott and Oxenham, 2008). This shows that lowering or raising a pitch pattern as a whole does not destroy it perceptual entity (Attneave and Olson, 1971). The perceptual entity of a pitch pattern, therefore, seems to be more related to the relationship among pitch events rather than the isolated pitch events themselves.

Despite the differences across domains, similarities might emerge even when one attempts to explore their differences. Although speech is not organised based on a set of specific intervallic relationships, music intervals might be also observed in speech. In Western music, certain types of intervals are associated with specific emotions. A 'major third' interval is normally identified as a 'happy' interval, whereas a 'minor third' interval relates to sad emotions in Western music.

It is of interest to explore if humans can perceive intervals in spoken utterances, despite the absence of fixed intervallic relationships in speech (Patel, 2008). There is evidence supporting existence of music intervals in speech but, at the same time, other researchers originally assuming the same hypothesis point to another relationship between speech and music intervals. Schreuder (2006) investigated the question of whether music intervals exist in speech employing two speakers to read two written passages. The first passage had a semantically happy content and the second passage was sad. Results showed that the speech of the former text included major third intervals, whereas the latter included minor third intervals. These two interval types are very common in singing and instrumental music in Western musical tradition. Ross et al. (2007) argue that one will intuitively look for the speech counterpart of musical intervals in intonation. That is, one can easily assume that it is meaningful to compare distances between pitch events in speech and music, wishing to discover some parallel between them, as in the case of Schreuder (2006). However, Ross et al. (2007) did not detect solid evidence for musical intervals in speech. Instead, they detected another parallel by analysing vowel nuclei spectra in adult American English participants. Their study found support for the argument that the human preference for specific tonal intervals in music arises from vowel properties in speech, as the ratio of the second to the first formant (for a more detailed description of formants see 2.1.1) in vowels often reflects one of the Western musical scale intervals.

As mentioned above, speech pitch organisation depends on relative rather than fixed relationships as is the case in music. However, both language and music are characterised by patterns of falling and rising pitch. According to Patel (2008), one cannot be sure whether these rising and falling pitch patterns are processed by the

same mechanisms in both domains. It seems that there are a number of different possibilities. For instance, pitch processing in language and music may be totally independent. Or, processing across domains might share some capacities. Alternatively, there might be a common mechanism for pitch contours for both domains, but an additional mechanism is likely to serve the specific aspects of pitch perception in music. That is, there may be a more refined and sophisticated mechanism, specific to the musical domain that might subserve processing of musical pitch stimuli. Zatorre and Baum (2012) argue in favour of two distinct mechanisms for pitch processing; a coarse perceptual mechanism, which can be used in both language and music, and a more fine-grained one, which is employed for pitch processing in music. Evidence for this comes from studies with individuals with musical impairments that are able to perform successfully on one type of pitch processing but not the other.

For example, in a study by Ayotte et al. (2002), individuals displaying congenital amusia were shown to perform normally on tasks examining perception of speech contours, whereas this was not the case for music pitch contours. Congenital amusia has been associated with significant difficulty in the detection of pitch changes, impairments of pitch direction detection, and also some problems with intonation contours extracted from speech (Patel et al., 2005). In a case study by Peretz et al. (2002), support for the argument that pitch processing in music and speech might depend on distinct networks is provided. In this study, a woman with congenital amusia was severely impaired at making pitch discriminations in music. Although her performance suggested a difficulty with pitch contour processing, she was shown to be able to monitor variations in speech on a task with internal and final pitch manipulations. Peretz and Hyde (2003) suggested that amusics are successful on

speech contour processing, because speech involves larger intervals. In western tonal music, on which most amusics have been tested, small intervals of one semitone occur very frequently. However, speech intonation contours include much larger intervals.

This hypothesis was tested by Patel et al. (2005) who studied seven congenitally amusic individuals. They assumed that if this difference between speech and music pitch was the reason of different performance in amusic participants, they would be able to discriminate speech contours when presented both with original sentences and delexicalised versions of these sentences. They found that their amusic participants did not succeed in perceiving these additional versions of speech utterances. Hence, researchers argued that demonstrations of successful performance on intonation tasks in previous research may not stem from the fact that speech utterances have a coarser pitch structure.

More recent research by Liu et al. (2012) also showed that it is not the organisation of pitch into coarse or fine-grained patterns that poses this difficulty for congenital amusic individuals. Rather, using stimuli with discrete and gliding pitches, they showed that it is the discrete versus gliding condition that brings about this impairment result in amusic populations. Hence, as discrete pitches are used in music rather than in speech, this deficit of congenital amusic individuals was previously thought to be specific to the domain of music.

Apart from the particular nature of pitch, the relationship of a given pitch with previous and succeeding pitch events also seems to influence auditory processing. Pitch representations in the human brain are often assumed to be the result of exposure to speech and music and the formation of pitch expectations in listeners

seems to apply both to linguistic and musical processes. The auditory system generally seems to be particularly effective in processing sequential auditory events (Lange, 2009). Exposure to Western tonal music may lead to the formation of expectations of future events that derive from its regularities with respect to, for example, tone repetition, pitch contour, or interval size (Marmel et al., 2008). That is, even the musically-uneducated listener can infer a successive event based on the patterns present in the music input. Therefore, if a musical passage gets interrupted and one is asked to guess the next interval missing from the recording, a correct guess is likely to be made. In the case of speech, one does not seem to infer successive events based on tonal relationships, but patterns governing interval size and relationships between different intervals appear to influence pitch perception in speech. As Schellenberg (2002) notes, children fail to have expectations for pitch direction changes after large intervals in melodies, because this pattern does not occur in the speech input they receive. That is, being exposed to linguistic input in which a change of pitch direction does not occur after a large interval, children develop perceptual habits that seem to influence their pitch expectations outside the speech domain, i.e., in music. This argument suggests an effect of speech processing on the processing of musical input on the level of pitch contours.

Pitch also seems to play an integral role in processing mechanisms related to emotion across the linguistic and the musical domain. Different speaker emotions may be perceived in intonational contours. Chen (2005) looked at the perception of surprise in British English and Dutch participants. Listeners were instructed to imagine themselves as the addressees of a number of utterances in their native language and indicate how surprised they thought the speaker of the sentences was. The focus of this experiment was to assess the effect of pitch-related cues on the

perception of this particular emotion. Results indicated a relatively similar performance across the two groups. For both British English and Dutch speakers, a higher pitch peak was judged to mark a larger degree of surprise. This is only an example of how a pitch contour might relate to the realisation and perception of a given emotion. Of course, one should expect significant variation in pitch-emotion relationships across cultures. It is, however, interesting when different cultures associate similar pitch contours with the expression of the same emotion, as this can raise further questions about the possible reasons why emotions are often closely related to specific pitch patterns.

To sum up, pitch accuracy is indispensable in music, whereas in speech, some degree of deviation from a given speech contour does not sound abnormal. Slight pitch manipulations applied to speech utterances are not perceivable in contrast to equivalent manipulations in music, possibly due to the fact that speech contours are not governed by tonality principles. One might not, thus, be able to perceive if a pitch manipulation has been applied when being exposed to an utterance where some pitches have been slightly manipulated. However, this issue is outside the scope of the present investigation. As mentioned above, the fact that slight pitch manipulations are easily perceivable pertains to another difference of the two faculties; namely the presence of tonality. In music, organisation of pitch is governed by tonal hierarchies. Even a slight deviation from a prescribed tone might result in a dissonant result. In speech, these principles do not apply, as it the relative direction of the change of pitch rather than a precise pitch that is instrumental for the formation of speech contours. These observations can be summarised in the distinction between two types of pitch processing mentioned in the literature; a

coarse type that relates to the pitch contours of speech and a fine-grained that accounts for complex pitch organisation of music in smaller intervals.

According to the above evidence, pitch appears to be an element that displays both similarities and differences across the domains of language and music. A great deal of research has focused on pitch in an attempt to explore the cognitive and neural substrates serving its processing. There are many unanswered questions concerning pitch in language and music. This parallel study may be extended to investigation of additional features beyond pitch or tone. The main scope of this study is the consideration of a new perspective that goes beyond the extensively studied organisation and perception of pitch across the domains.

1.4 The language-music relationship and theories of cognition

Establishing links among different cognitive functions and also links between disruptions of these functions and brain regions has proved particularly challenging. Moreover, it is hard to determine whether the brain is equipped with innate linguistic and musical 'devices' or it adapts to environmental input later in life through exposure. This section is devoted to theoretical accounts of higher cognitive functions and their relevance to the language-music relationship.

Studying brain-damaged individuals, researchers have attempted to explore performance on linguistic and musical processing and their possible neural relationship. A given deficit or a combination of deficits might be often detected in studies with patients with large brain lesions. Music disorders in particular often appear as a result of extensive lesions. Acquired amusia seems to result from damage to extensive areas bilaterally but may also arise from more restricted lesions either in

the left or in the right hemisphere (see, for example, Patel et al., 1998 and Nicholson et al., 2003). It is, therefore, hard to use such evidence to construe direct localisation-function links, as a large source of data comes from studies with patients either suffering from large lesions or displaying similar deficits caused by injuries to several different regions of the brain.

Modularity constitutes a central theoretical concept in cognitive science. The concept of 'modularity' has been employed by researchers in the field as a framework for interpreting the relationship between linguistic and musical impairments. According to Fodor (1983), the mind consists of modules which are characterised by a series of features; domain-specificity, rapidity of operation, automaticity, informational encapsulation, innateness, and neural specificity. In a further elaboration of this concept by Karmiloff-Smith, a module can be defined as an information-processing unit encompassing knowledge in a domain, such as language or arithmetic and the relevant computational force (Karmiloff-Smith, 1995). Each domain can be further subdivided in smaller subdomains. The exploration of the relationship between language and music can be subject to the different conceptualisations of modularity (McMullen and Saffran, 2004). In the comparative study of language and music, the question arises as to whether these two faculties are served by distinct mental modules or their mental processing shares some components. Depending, thus, on the conceptualisation of modularity, one can look, for example, at issues of 'neural specificity' or 'innateness' concerning these two cognitive domains. In a version of modularity that includes epigenetic factors, Karmiloff-Smith (1995) suggests that, despite the native predispositions and the other 'Fodorian' features that might exist in the human brain, early development in infancy might be the product of constant

interaction with the environment. Thus, this milder version, 'modularisation,' seems to compromise cognitive predisposition and brain plasticity.

There has been some debate as to whether music processing is modular or it is linked to linguistic processing though general cognitive functions. Peretz and Coltheart (2003) argue in favour of a modular architecture of music processing. They suggest that the fact that many individuals whose brain damage has selectively affected musical abilities while sparing other abilities, such as speech or environmental sound recognition, implies the existence of a music-processing module, that is, a processing system specific to music. According to the principle mentioned above, the music module is further divided into additional modules. In contrast to the idea that music is modular and independent from other cognitive functions, Schuppert et al. (2000) argue that music perception is not based on specific music processing networks, as several cognitive functions (working memory, phonological loop and executive functions) also seem to contribute to music processing. It is of interest to cognitive neuroscience, though, to investigate whether the shared processing mechanisms of linguistic and musical functions are entirely due to underlying general cognitive abilities, such as attention or executive function, or whether there is a more significant direct link between the two domains.

Data emerging from neuropsychological and brain imaging studies have provided contradictory results, as those involving individuals with brain damage have pointed to distinct processing dissociations while studies involving healthy participants have demonstrated shared processing across domains. In order to reconcile these conflicting sources of evidence, Patel (2012) proposed a relevant theoretical framework. This framework, namely the 'resource-sharing framework', suggests that knowledge can be domain-specific but, as processing takes place, shared neural

networks will be employed in both domains. According to this framework, there might be domain-specific representations in long-term memory which can be selectively compromised, giving rise to dissociations often reported in the literature. If these hypotheses are true, neuropsychological dissociations do not prove the existence of distinct cognitive domains. However, given the fact that this is a theoretical assumption and not a data-driven conclusion, one can still argue that a dissociation observed might pertain to distinct processing until the 'resource-sharing framework' is supported by solid empirical data. As far as whether domain specific cognitive abilities are innate or acquired through experience, the relevant explanation does not necessarily need to be entirely modular. Studying cognitive abilities in children suggests a more dynamic process between the human mind and the input it receives from its environment (Karmiloff-Smith, 1995).

In brief, a major hurdle in the 'decoding' of function-localisation relationships when language and music are studied in parallel is the fact that neuropsychological research data often come from patients suffering from damage to large areas of the cerebral cortex. Acquired amusia does not typically arise from small lesions, or with a consistent localization or lateralization in contrast to some patterns found in acquired aphasia. From a more theoretical angle, the theory of 'modularity' suggests that the mind has a number of modules that are independent and are devoted to a specific type of processing. 'Translating' this theoretical idea to the comparative study of language and music, researchers are interested in determining whether there is evidence for independent modular mechanisms for language and music or if they have shared neural resources. Empirical results have been ambiguous. Of course, it is often hard to rule out general cognition processes that are likely to be involved in both language and music. When one studies language and music, *de facto* cognition

comes into play. It is, therefore, challenging for researchers to design appropriate tools in order to examine genuine language-music relationships in a cognitive framework appreciating –to the greatest possible extent– the role of general cognitive abilities in linguistic and musical processing.

1.5 Differences between individuals with and without musical experience

Studying musicians' cognitive processing suggests that music training can bring about plastic changes in the human brain. This can be more clearly shown when musicians and non-musicians are studied in parallel. Research suggests that musicians' brains appear to adapt physiologically to the music training the receive (Zatorre et al., 2007). However, the exact plastic changes are not yet very clear. Musical skills have not been matched with specific circumscribed brain areas and there is no established anatomy for musical processing (Gaser and Schlaug, 2003). A closer look at brain structure and music cognitive function suggests that additional systematic research can result in a possible mapping of cognitive functions in the future.

Schlaug et al. (1995) investigated whether musical training affects intrahemispheric relationships in individuals with this kind of experience. The corpus callosum, a structure of fibres connecting the two cerebral hemispheres, was found to have a different morphology in musicians in their study. Musicians were shown to have a larger anterior part of the corpus callosum in contrast to non-musicians. This study raises the issue of plasticity differences due to musical training in early childhood, as these researchers looked at musicians with training that took place before the seventh year of life. As these differences were observed in young individuals with early

music training, the study implies that a maturation period in the early years of life is likely to exist. With respect to the relationship between memory for music and music training, hemispheric differences also appear. For example, Ayotte et al. (2000) argue that areas of the left hemisphere appear to be more strongly related to long-term music representations, whereas areas of the right hemisphere seem to provide access to these representations. This might imply that training -which contributes to the formation of these long-term representations- might depend more on left-hemisphere structures.

In comparison to non-musicians, musicians seem to display a different cognitive response when they are exposed to pitch stimuli. Perception of frequencies is not a process involving the same localisation for the whole range of different frequencies (see 1.6). Tonotopy, a general function of the auditory system, shows a frequency-dependent localisation in the cochlea and, at later stages, in the central auditory pathway (Schonwiesner et al. 2002). This tonotopical representation means that when one is exposed to an acoustic stimulus, a different frequency will activate a different region of the cochlea. The same principle is the case for the cerebral cortex. Music training has been argued to have an effect on how pitch stimuli activate different parts of the cortex. Early investigations in differences in pitch processing in musically educated participants and musically uneducated controls show anatomical differences in the auditory cortex. More specifically, Pantev et al. (1998) found that musicians that were exposed to piano tones displayed more extensive areas of activation, according to the results of functional magnetic source imaging that researchers used. This difference in representation was even larger in musicians whose musical training started at a younger age compared to individuals that had never played a musical instrument.

Another question of pivotal importance in this area has been whether musical training can create brain plasticity extending beyond the musical domain. As already discussed, perception of contours differs from perception of specific intervals. Thus, although pitch patterns in terms of direction are perceived more easily, the accuracy needed in order to recognise specific intervallic relationships appears far more demanding. Researchers in music cognition have looked at the effect of music training on pitch processing not only in music but also in language. Besson et al. (2007) used electrophysiological measurements to test the perception of pitch in musicians and non-musicians, using both musical and linguistic stimuli. In these stimuli the final note or word was manipulated so that the stimuli would become incongruous. Moreover, there was variation in the amount of incongruity that was created so that some melodies and utterances' ending sounded very incongruous and in some other cases slightly incongruous. Results showed that, in cases where incongruities were easy to detect, musicians and non-musicians performed similarly. However, for slight incongruities (such as 1/5 of tone manipulations) musicians performed much higher on both the musical and also the linguistic tasks. These results suggest that musical training can have an effect on pitch perception not only in the musical but also in the linguistic domain. In other words, this skill seems to be transferable to the linguistic domain without speech perception training.

The pitch acuity benefit acquired by music training seems to extend to languages that one does not speak. Marques et al. (2007), a follow-up study similar to Besson et al. (2007), used the same test in order to test perception of final-word incongruity in a language that participants did not speak. French participants, musicians and non-musicians, were exposed to sentences in an unknown language—Portuguese, in which similar manipulations to those in Besson et al. (2007) were applied.

According to results, slight pitch incongruities were only perceivable by musicians but not non-musicians. This means that slight sentence contour manipulations can be perceived when one is musically trained even when it comes to a foreign language of which they do not have any knowledge. As music pitch perception involves processing of much smaller intervals compared to speech, musicians' ability to perceive fine-grained manipulations in music might be transferred to the speech domain. Hence, this transfer effect does not seem to be linked to the lexical aspect of speech and it is not surprising that musicians are sensitive to this kind of manipulations in foreign languages.

The results of these studies and other studies on similar topics are likely to be biased in favour of a critical contribution of music training to musical intelligence and cognitive performance in general. However, if one takes into consideration that music ability is not a single cognitive entity but rather is comprised of multiple sub-components (Peretz and Coltheart, 2003), there may be multiple effects of music training to a range of music abilities that have not yet received scientific attention. The question arises as to whether the range of music abilities can be divided into categories concerning skills that are highly dependent on specific or early life training and those that can develop in any human being after simple exposure to musical input. Investigation of this issue might also lead to a clearer picture about which musical abilities are specific to the music domain and which might relate to other cognitive domains or to general cognitive abilities.

As musicianship offers the opportunity to study individuals with a long period of specific training resulting in processing differences with non-musicians (Bangert et al., 2006), this area is likely to produce even more fruitful research in the future. The present study will further extend the findings from the existing evidence coming

from musicians and non-musicians literature with investigations focusing on those musical abilities that may emerge as a result of simple exposure rather than long-term musical training.

1.6 Disruptions of speech and music processing

Before embarking on the consideration of impaired speech and musical processing, it appears meaningful to start from a short description of the processing of sound in general in the healthy auditory system. Sound, defined in physical terms as the pressure waves produced by vibrating air molecules, is transmitted through the external and middle ear to the cochlea of the inner ear where it is transformed into neural impulses (Purves et al., 2004). Processing of complex sound such as speech and music relates to its detection in at least one of four possible dimensions, namely, frequency, time, amplitude and space (Griffiths et al., 1999). Since higher level auditory processing and its impairment is of specific interest in the present study, processes involving the more peripheral aspects of the auditory system will not be reviewed here. Although considerable processing takes place in subcortical structures (Zatorre et al., 1992), the auditory cortex appears of paramount importance for the discrimination and localisation of acoustic stimuli as well as auditory learning and memory and consists of multiple areas of specific anatomical and physiological characteristics (Budinger, 2005). Hence, there are distinct areas in the cortex subserving different stages of auditory processing. All mammals have cortical regions dealing with the first phase of sound processing in the cortex, termed primary auditory cortex (Kaas and Hackett, 2005). The frequencies in the acoustic input one receives are mapped spatially both at the cochlea and later in the auditory cortex. This spatial organisation is referred to as 'tonotopy'. More specifically, areas

involving different neurons relate to different frequencies and, therefore, frequencies close to each other are represented by adjacent regions (Purves et al., 2004).

As far as language is concerned, it is important to note that the divisions of the auditory cortex highlight the difference between sound in general and speech. Primary auditory regions can be activated when one is exposed to noise bursts but not with speech, even if those noise bursts have similar acoustic properties with speech (Zatorre et al., 1992). Further support in favour of this distinction comes from studying congenital deaf individuals. Nishimura et al. (1999) investigated how auditory areas process information in congenitally deaf participants fitted with cochlear implants with the use of neuroimaging techniques. As pre-lingual deaf individuals are able to hear some sounds with the use their implants, they compared such participants' processing of sounds and of sign language (Japanese Sign Language). Results showed that the primary auditory cortex was activated when these participants were exposed to sounds and the secondary auditory cortex processed sign language. Their conclusion, therefore, was that visual gestural linguistic input can activate those areas devoted to spoken language in normally hearing people but not areas of the primary auditory cortex. These findings emphasise both the fact that sign language   is a language processed similarly to spoken languages and also that the primary auditory cortex is not involved in the complex processing of linguistic stimuli. Different responses of the auditory cortex according to the specific type of a stimulus are also observed for music. For instance, the auditory cortex seems to process differently contours from specific intervals. It has been suggested that, in brain-damaged populations, processing of intervals might be impaired without affecting processing of contours, but, in general, these mechanisms are poorly understood (McDermott and Oxenham, 2008).

Given the fact that pathological voices are often part of the input in a human linguistic environment, it is of interest to look at how listeners perceive these voices and if they can do discreet discriminations of different acoustic qualities inherent in them. As it will be discussed later in 2.1.3, voice quality refers to a series of features such as pitch, loudness, or timbre that contribute to the characteristic colour of one voice that distinguishes it from another. In other words, it is a combination of features that enable us to discriminate voices of familiar and unfamiliar people. Dysphonia is an example of pathological speech. It is a disorder marked by altered voice quality in which problems with pitch, loudness and excessive vocal effort can lead to communication problems and quality of life disadvantages (Schwartz et al., 2009). Dysphonic speech can sound breathy and more quiet, as it happens sometimes with non-pathological speech. A breathy voice is generated when the vocal folds have an incomplete closure along their length and in English-speaking countries it is sometimes associated with 'attractiveness' in female voices used for advertisements (Ogden, 2009). As the categories pertaining to different aspects of speech (e.g., linguistic intonation, emotional intonation, pathological speech features) are not exhaustive, attempting to limit this multi-dimensionality into specific labels rather than dynamic descriptions might be problematic.

Eadie and Doyle (2005) studied perception of dysphonic voices in a connected speech context. The authors argue that connected speech appears to be a more relevant stimulus in comparison with isolated vowels, as human communication is not conducted via the use of isolated vowels that might lack essential acoustic information. Therefore, perceptual judgements related to voice quality and pathological characteristics can be studied in a more genuine way by connected speech stimuli. Twelve participants were asked to judge and rate pleasantness, being

exposed to normal and dysphonic speech that was acquired by readings of standard text testing connected speech output. Hence, all listeners were exposed to 24 dysphonic samples (from 24 different speakers with dysphonia) and another 6 samples of non-pathological speech. At the same time, acoustic analysis was conducted for the same samples. Based on their findings from both the listeners' responses and the acoustic analysis, the researchers note that one cannot draw clear parallels between perceptual judgements of voice pleasantness and their corresponding acoustic measures.

An important distinction should be highlighted between hearing and high auditory functions at this point. Perception of language and music is not necessarily preserved in the presence of normal hearing. That is, an individual's hearing might be completely intact, but their perception of linguistic or musical patterns can be severely compromised. One type of disordered processing of speech, language, or environmental sounds despite normal hearing is referred to as 'auditory agnosia' and is further divided into word deafness, amusia, and environmental sound agnosia depending on the impaired faculty (Griffiths et al., 1999). Patients suffering from word-deafness are unable to understand spoken language and, in the case of pure word-deafness, a patient displays impaired speech comprehension, despite their intact expressive language ability, the production of written language, and the ability to comprehend written language (Gutschalk et al., 2015). Amusia can be either acquired or congenital. Acquired amusia can be the result of brain damage, whereas congenital amusia is a developmental musical deficit that cannot be attributed to brain injury, hearing, cognitive or social etiologies (Peretz et al., 2002). However, the distinction between agnosia affecting music and agnosia affecting speech seems not well defined.

As it was shown in 1.3, the study of individuals with congenital amusia has contributed to a large extent to the study of pitch processing. Congenital amusics have been shown to have compromised pitch processing abilities (Ayotte et al., 2002; Foxton et al., 2004; Hutchins et al., 2010; Hyde & Peretz, 2004; Jiang et al., 2011; Liu et al., 2010; Liu et al., 2012) which have been often attributed to impaired working memory (Albouy et al., 2013; Gosselin et al., 2009; Tillmann et al., 2009; Williamson and Stewart, 2010). The question arises as to whether this pitch deficit is also present when it comes to linguistic processing. In some studies, this has proven to be the case. For example, two studies presented in Patel et al. (2008) reveal an impairment of pitch processing in congenital amusics across domains. Researchers presented congenital amusic participants with stimuli of identical lexical content but with different intonation contours. In most cases, the last syllable or word of an utterance was produced as an upward or downward pitch glide, depending on whether the stimulus was a question or a statement. An additional condition included nonlinguistic tone sequences by replacing each syllable with a tone. Comparing the results across the two task conditions, Patel et al. (2008) suggested that some amusics are not be able to discriminate statements from questions due to impairments in perceiving final pitch directions. Liu et al. (2010) also found that the pitch impairment of individuals with congenital amusia goes beyond the music domain, affecting speech intonation processing. Their congenitally amusic participants manifested compromised performance on processing statements and questions, failing to detect the pitch direction differences in the final word of the utterances to which they were exposed. In contrast to previous studies (e.g., Ayotte et al., 2002, Patel et al., 2005), Liu et al. (2010) used smaller pitch contrasts (but valid in the speech context) in their stimuli. This methodological manipulation

revealed that congenital amusics do have problems with speech intonation apart from music contours. It appears that congenital amusia does not affect individuals in identical ways. In Patel et al. (2008), about one third of amusic participants were impaired at intonation processing in formal testing, whereas the rest were not. However, this might reflect methodological choices rather than definitive assessments of processing. The amount of research evidence that has been gathered on congenital amusics can be usefully expanded by consideration of evidence from acquired music processing deficits to gain a deeper understanding of the music-language relationship.

As mentioned above, division of auditory agnosia into large categories might not be sufficient. Breaking down auditory perception into specific subcomponents to be tested reveals a more complex picture than a speech-music dissociation. In a ground-breaking study on auditory agnosia, Peretz et al. (1994) investigated whether perception of linguistic and musical acoustic features are interconnected. They tested two patients with bilateral damage in the auditory cortices on a number of tasks. The following detailed description of the tasks and the patients' performance aims to highlight the multiple aspects of music perception and the complexity in the pattern of performances. When the two participants were presented with a task on recognition of familiar melodies, they both displayed an impaired performance. One of them, however, performed better in terms of retrieval of information when provided with some lexical information concerning the melody. A more complex picture emerged with respect to discrimination and recognition of unfamiliar tunes. Both participants were successful in comparing pairs of stimuli differing in rhythm, but they were impaired at comparing melodies differing in pitch events. However, they could successfully compare the relative pitch of two single notes. Their

recognition of environmental sounds was also preserved. However, their perception of prosody was not without problems. The first participant had intact emotional prosody perception but her linguistic prosody perception was below the range of controls. The second participant's perception of linguistic prosody was within the normal range but when it came to emotional prosody perception, he often confused sadness with joy in several pairs of stimuli. The patients could not identify famous public figures based on their voice but they could discriminate unfamiliar voices based on gender. One of them could discriminate voices based on the presence of foreign language accent but not so the other. Despite the common postulation that pitch contours in speech and music are processed similarly because of their shared acoustic properties, the performance of the participants in this study shows that normal processing of contours in one domain does not ensure also unimpaired processing in the other domain. An additional interesting point that emerges from these findings is that, although it is not discussed in their study, might have to do with different aspects of pitch processing. The participants of their study had an impaired ability to discriminate between melodies that differed in terms of pitch. At the same time, they could discriminate unfamiliar voices based on gender which might imply that they relied on pitch perception mechanisms in order to achieve it. A similar issue is treated in the discussion of the present study, highlighting the importance of comparing pitch perceptual abilities of different kinds.

Processing of speech prosody (for more details, see 2.1.3) and processing of music pitch contours have been investigated through lesions studies that reveal a critical involvement of the right hemisphere. Despite several accounts of the contribution of the left hemisphere to this type of processing both for speech and music (e.g., Pell and Baum, 1997, Patel et al., 1998), the focus here is on the music studies that

provide evidence on the role of the right hemisphere in pitch processing through the study of brain-damaged patients. The evidence on the role of the right hemisphere in speech processing is examined in parallel.

In general, the right auditory cortex appears to be dominant in identifying pitch (Zatorre, 1998). It is also thought to be dominant in recognising melodies (Stein and Stoodley, 2006). Individuals with lesions in the right hemisphere are shown to have difficulty in perceiving pitch patterns in melodic stimuli (Liegeois-Chauvel et al., 1998). In a more specific functional neuroanatomical analysis, Zatorre et al. (1994) found that: superior temporal areas are involved in perceptual analysis of melodic contours; pitch comparisons appear to be conducted in right prefrontal areas; and retention of pitch patterns is likely to be a result of processing in both right temporal and frontal areas. Hence, damage to these areas is likely to compromise perception but also retrieval of pitch patterns in music. A more recent study (Warrier and Zatorre, 2004) with right temporal lobe patients is in line with this evidence, highlighting, at the same time, the hypothesis that compromised memory for contour might account for impaired performance in patients with right hemisphere lesions.

The role of the right hemisphere seems to be even more critical when a pitch task becomes more demanding. More specifically, Johnsrude et al. (2000) found that patients with temporal lobe excisions in the right hemisphere were impaired at judging the direction of pitch change, in contrast to patients that displayed a similar lesion in the left hemisphere. The right hemisphere also seems to play an important role in processing speech contours. Meyer et al. (2002) suggest that areas in the right temporal lobe might be also dominant in the processing of pitch patterns in speech (pitch contours). One can, therefore, suggest that lesions in areas of the right

temporal lobe are associated with impairments in perceiving pitch contours across domains.

The study by Nicholson et al. (2003) is the only study, to the author's knowledge, that has examined perception of linguistic and musical prosodic features in a patient with right hemisphere damage (the same patient also participated earlier in Steinke et al., 2001). This patient suffered from a right fronto-parietal stroke and displayed compromised ability on prosodic and musical discrimination tasks, in the presence of intact segmental speech perception. These findings were obtained by using the tasks designed by Patel et al. (1998) which are described in section 3.3. The researchers suggested that this evidence points to shared prosodic and melodic –at least partially– neural substrates in the brain.

Due to the nature of the experiments conducted for the purposes of the present study, this section has reviewed previous research evidence related to compromised perception of acoustic stimuli rather than their production. The study of auditory processing over recent decades has demonstrated that the auditory cortex plays a significant role in discriminating and localising sounds as well as facilitating learning and consolidating auditory memories. The auditory cortex displays anatomical and physiological differences that seem to participate in different stages of processing and in the analysis of different types of stimuli. Sounds other than speech are processed differently from actual speech, even when they bear similar acoustic properties. The secondary auditory cortex appears to be active in processing complex linguistic stimuli, while the primary auditory cortex appears to be primarily responsible for more simple acoustic events. Early investigations with brain-damaged participants revealed that the processing of music tunes differs significantly from the processing of lyrics accompanying those tunes. As far as the perception of

speech and music pitch contours is concerned, studies with congenital amusics and those with acquired amusia as a result of brain damage do not provide converging results, as many associations and dissociations of processing across domains have been found in different individuals.

With respect to the question of lateralisation of function, the right hemisphere has been shown to contribute significantly to pitch processing. However, involvement of left hemisphere areas has been also reported. Hence, the current picture is complex. The last decades have revealed a multitude of new discoveries in the auditory domain and the relationship between speech and music. However, studying related speech and music processing in individuals with impaired cognitive functions has not produced a coherent picture. Thus, more detailed study of neurological deficits in music and speech processing appears of paramount importance in the appreciation of these cognitive abilities and their relationship.

2. Defining a new prosodic aspect of language and music

Starting from an account of work related to prosody in language and music, this chapter demonstrates the limits of past research which has focused on a narrow set of variables and expands into the original features of the present study. Previous work on language and music perception refers to various prosodic features and their possible connections across the domains. However, a gap in research hitherto conducted is described and analysed here.It highlights the need to consider an additional prosodic aspect which is not covered by previous investigations. This new perceptual aspect is presented as a dynamic combination of acoustic cues that

contribute to the aesthetic appreciation of speech and music streams and will be referred to as 'expressiveness'. The term is defined by gradually eliminating all the features that pertain to other aspects of prosody identified in past research.

2.1 A note on segmental and suprasegmental aspects of speech

This section presents some elements of segmental and suprasegmental phonology in order to introduce phonological functions in later sections. Terms such as, 'formants', 'phonemes', 'frequency', 'prosody', and 'pitch' along with other relevant concepts are discussed.

In order to appreciate the difference between descriptions of briefer phonetic events and those occurring over longer durations, it is meaningful to start from the smallest levels of phonological description and elements of prosodic analysis that appear more relevant to long musical streams. As the emphasis of this dissertation is on features of language and music extending over streams of seconds rather than milliseconds, consideration of segmental speech properties will be limited, only to serve the role of juxtaposing short with long acoustic streams.

2.1.1 Formants

Speech production is a multifaceted process. As mentioned in 1.6, sound is the result of waves produced by vibrating air molecules. The difference between mere production of sound by vibration and production of speech is that the latter results from a more controlled effort in which the speed of vibration and tension of the vocal folds is combined with the size and shape of the vocal tract (Ball and Rahilly, 1999). Identification of speech sounds depends upon the listener's perception of

formants. Crystal (2008) gives a detailed definition of the term. A formant can be

understood as the concentration of acoustic energy that depends on the vibration of

air from the lungs in the vocal tract consequent to changes in its shape. Hence, each

configuration of the vocal tract corresponds to different formants. The positions of

the articulators, such as the jaw, tongue or the lips contribute to the production of a

given formant (Ogden, 2009). That is, some frequencies resonate more than others

depending on the final shape of the vocal tract and the articulators and this results in

the production of a vowel instead of another. It is important to note that formants do

not solely provide acoustic information on vowels. Formant transitions show a shift

of articulatory changes in speech between consonants and vowels, in which the vocal

tract changes from a shape imposing constriction of air into an unobstructed resonant

shape (Clark et al., 2000).

In an alternative description of formants, Purves et al. (2004) draw a parallel

between the vocal system and musical instruments: as the reed of a clarinet and the

whole instrument structure, the vocal system has a similar 'reed', the vocal folds, and

space between the vocal folds and the lips. Hence, the pathway between the vocal

folds and the lips or the nostrils determines speech sounds as the structure of the

musical instrument determines musical sounds. As phonemes produced by different

individuals do not have identical acoustic properties due to anatomical differences in

body size, notes of the same pitch will have a different quality when they are

executed by different instruments e.g., piccolo and bassoon (Ball and Rahilly, 1999).

The fact that speakers have different vocal tract lengths and shapes and also the fact

that while articulating, there is transition from one target configuration to another

makes it impossible to define speech sounds in absolute terms, as these vary across

speakers (Blumstein and Stevens, 1981). This phenomenon known as 'lack of

invariance' suggests that listeners have mechanisms to encode speech sounds despite the fact that those are not identically produced by all speakers or speech contexts.

2.1.2 Segmentals

Although the linguistic identity of speech sounds cannot be determined in an absolute fashion, it is still possible for the listener to identify and discriminate between different speech sounds. Different speech sounds, as these are perceived by listeners, are called 'phonemes'. Defining phonemes becomes easier if one thinks of the lexical context in which speech sounds occur; a phoneme is the minimal unit in the sound system that results in semantic contrast (Crystal, 2008). That is, in order to determine if two phonemes differ in a given language, one can put them in lexical context and compare them. If this process leads to two existing words in a given language, the difference in meaning indicates the existence of two different phonemes (e.g. 'bin' and 'pin' in English). A crude division of phonemes in two categories, consonants and vowels, can be explained by a significant difference in their production. Consonants are produced by obstructing an air-stream in the pharynx or in the vocal tract but vowels are produced without any obstruction and, in comparison to consonants, they bear larger amount of vocal effects such as loudness or pitch (Skandera and Burleigh, 2005). There are additional subdivisions of vowels and consonants but this is beyond the scope of this study. These subdivisions correspond to natural classes that are organised according to features which can be described in a systematic way in the sound system (Blumstein and Stevens, 1981).

2.1.3 Suprasegmentals

Segmental features of speech are produced and perceived in a different time frame from other phonetic features, referred to as 'suprasegmentals'. The term itself reveals that these features have to do with vocal effects that extend over more than a single segment (Crystal, 2008). When one talks to a baby or an animal, they often use a combination of suprasegmental features, as for example, changes in voice quality or higher pitch register (Ogden, 2009). This means that these features have a role which can be viewed beyond the lexical aspect of speech. Despite the fact that these features do not convey literal meaning *per se*, their extralinguistic meanings may contribute a great deal to communication. It is therefore worth examining these features individually in order to appreciate how these are described acoustically, how they are perceived auditorily and, in later sections, how these can relate to acoustic streams of music.

Prosody, another term for suprasegmentals, encompasses timing, frequency, amplitude, pausing, and voice quality. These are variables that mark all parts of a spoken utterance and, according to the existing literature, contribute to the formation of three main types of realisations: linguistic, pragmatic, and emotional prosody. In general, researchers use the term 'prosody' to either refer to an abstract definition without looking at any specific prosodic components or to examine closely the above features of timing, frequency, amplitude etc. (Cutler et al., 1997).

Loudness as a suprasegmental feature can contribute to disambiguation of meaning (in the case of heteronyms, as for example in the word contrast which can be either a noun or a verb) and communication of emotions (Skandera and Burleigh, 2005). Loudness can be relative in terms of perception. That is, other factors, co-occurring with loudness, can affect how loudness is perceived. For example, the pitch of an acoustic stimulus can affect perception of its loudness (Clark et al., 2000).

Length is also used to describe speech signals at the suprasegmental level. It refers to the physical duration of a sound or an utterance and it normally differs from duration that pertains to the time devoted to the articulation of a sound or a syllable at the segmental level (Crystal, 2008). Tempo refers to the speech rate of an utterance and differences at this level have a different function compared to segmental manipulations of duration. An important point is that differences in tempo do not produce differences in meaning equivalent to the differences that duration can bring about at the word level (Lehiste, 1970).

Another element that is taken into consideration for prosodic investigations of speech is 'voice quality'. In simple terms, voice quality is defined as the difference in 'colour' that one perceives among different voices and resembles the difference that one perceives when they are exposed to two identical notes played in equal loudness by two different instruments (Skandera and Burleigh, 2005). In this sense it can be alternatively called 'timbre'. A breathy voice and a harsh voice can be perceived as different even when they display the same fundamental frequency and loudness in the same sense that piano and violin timbre differences are perceived. However, in a broader definition of voice quality, the term refers to a series of features. It can refer to someone's rate of speech, pitch height, loudness and timbre (Crystal, 2008) rather than timbre exclusively. In this second definition, the term seems to encompass all these features that one can have at their disposal in order to identify a speaker.

Pitch is an important auditory feature which has received a lot of scientific interest and has also been the focus of research in many studies in the comparative framework of language and music. Pitch is the percept that relates to the acoustic

feature of frequency. The frequency of the vibration of the vocal folds determines the auditory result that one perceives (Skandera and Burleigh, 2005). That is, a fast vibration of the vocal folds results in a higher pitch, whereas a slower vibration results in a lower pitch. In contrast to most sounds that surround us and which are called complex sounds, there are also pure tones that differ from complex sounds in that they contain only one frequency (Griffiths et al., 1999). By contrast, the sound resulting from the vibration of the vocal folds is an example of complex sound with many associated harmonics (Lehiste, 1970). Pitch and fundamental frequency do not relate in a linear way. Fundamental frequency is defined as the frequency of a periodic (regularly repeating) sound that corresponds to the lowest mode of vibration and harmonics as whole number multiples of this frequency (Zatorre et al., 2007). Ogden (2009) explains that the relationship between fundamental frequency and its percept, pitch, is of logarithmic nature. More specifically, if this relationship was absolute, then the difference between 100 Hz and 200 Hz would be equal to the difference between 200 Hz and 300 Hz. Rather, fundamental frequency and the stimuli we perceive relate in a proportional fashion. That is, the difference between 100 Hz and 200 Hz is similar to that between 200 Hz and 400 Hz, meaning that the two stimuli have the same difference in proportion; 1:2 in both cases.

In simple auditory terms, pitch can be defined as the type of sensation scaled from 'low' to 'high' (Crystal, 2008). As also noted in 1.3, there is a distinction between absolute and relative pitch. Relative pitch requires the listener to abstract intervallic relationships (De Cheveigne, 2005). That is, the listener has a point of reference at their disposal and they base their judgement on the relationship between two notes rather than the identification of a note out of melodic context. Absolute pitch or perfect pitch means that the listener is able to identify a specific note without a point

of reference, that is, to name an isolated note. The use of pitch in prosodic functions is discussed later in this chapter.

2.2 Prosodic processing and lateralisation

Work on the perception of segmental acoustic features has been conducted in studies looking at several properties such as pitch (e.g., Klein et al., 2001, Luo et al., 2006, Xu et al., 2006). Studies on pitch have focused on different properties and time domains of pitch. It is here argued that the mechanisms devoted to pitch perception at the segmental level can differ significantly at the suprasegmental level. For example, lateralisation of pitch patterns at the segmental level can be different from pitch patterns included in a prosodic context. This section will discuss processing differences occurring as a result of different prosodic types, different nature of stimuli, different time domains, as well as issues concerning the integrated processing of such stimuli by the two cerebral hemispheres.

As mentioned above, it is important to note that results stemming from perception studies do not necessarily apply to production studies and vice versa. Rather, perception and production have been investigated independently. For instance, perception and production of emotional prosody appear to pertain to different processes subserved by distinct neuroanatomical systems (Borod, 1992). Focal damage to the right cerebral hemisphere can affect comprehension, repetition, and production of emotional prosody, depending on the exact location of the damage (Ross et al., 2001). Hence, one patient with a right hemisphere lesion might display perception impairments whereas a different right hemisphere-damaged individual might have poor performance on a production task. This can be the case even when it

comes to tasks focusing on the same emotion. Data from Borod et al. (1986) show that the inability of right hemisphere-damaged patients to produce a specific emotion was not found to relate to their inability to identify this emotion. One can, therefore, safely assume the existence of different processing mechanisms; this justifies studying production and perception of prosody separately. Examining prosodic processing in populations with deficits consequent to brain damage can shed some light on how this processing is likely to take place in the healthy brain.

From an anatomical perspective, the auditory cortices of the two hemispheres have been found to display asymmetries, with the left hemisphere having a larger volume of white matter (Penhune et al., 1996). Differences are not only anatomical but also functional. In contrast to the role of the left hemisphere in processing segmental information, the right hemisphere seems to have an advantage for tonal processing, according to a large amount of empirical data (Zatorre et al., 2007). The neural responses of the two cerebral hemispheres also appear to differ at the level of temporal processing of the speech stimulus. In order to appreciate the contribution of the left and right hemispheres, it is worth looking at the different 'decoding' strategies they employ in this type of speech processing. Zatorre et al. (2002) argue that the auditory cortices of the two hemispheres display some specialisation in acoustic processing. More specifically, the left auditory cortex appears to have better temporal resolution whereas the right appears to 'outperform' the left in terms of spectral resolution. The authors suggest that the good temporal resolution of the left hemisphere might account for its dominance in speech processing. This initial advantage for speech decoding, stemming from a single processing principle, might account for the hemispheric differences of speech perception.

An issue related to timing in processing is portrayed in the 'asymmetric sampling in time' (AST) hypothesis (Poeppel, 2003). According to this hypothesis, the speech signal is processed by both hemispheres but it includes various time scales (pertaining to formant transitions, syllables, and intonation contours). Although both hemispheres participate in the processing of acoustic stimuli, these are elaborated in an asymmetric fashion in terms of time. That is, the left auditory cortex mainly encodes short auditory signals, while homologue areas in the right hemisphere encode longer acoustic chunks.

In order to test the above hypothesis, Baum and Dwivedi (2003) worked with unilaterally brain-damaged patients and tested their perception of prosodic information employed for syntactic disambiguation, which demands processing of long chunks of information. They used stimuli which consisted of sentence pairs that began identically but the syntactic structure of their second half differed; it had an either minimal or non-minimal attachment ending. That is, one sentence stimulus included a verb structure that corresponded to the simplest possible interpretation ('considered the offer') and another sentence with a verb that corresponded to a more complex interpretation ('considered the offer was an insult') for which the listener needed additional time to make, given that they needed to have access to the next chunk of the stimulus. The two sentences also differed prosodically, having different temporal and intonational cues. For the sentences with a more complex syntactical structure, right hemisphere-damaged patients showed a slower reaction time but, in contrast to controls, the presence of incongruent prosody did not have an effect on the reaction time when processing either simpler or more complex syntactical structures. The left hemisphere-damaged patients that took part in the study could perceive prosody but, quite surprisingly, their reaction time when exposed to

incongruent prosodic markers was faster than the examples with congruent prosody. The performance of the right hemisphere-damaged participants corroborates the hypothesis that the right hemisphere processes prosodic information of a longer time span.

Absolute notions of lateralisation of function, however, might constitute an oversimplified account of cognitive processing. The consideration of degree of complexity inherent both in the acoustic stimuli and the mechanisms by which the two hemispheres process such stimuli can reveal a more complex picture of prosodic processing. Friederici and Alter (2004) suggest that lateralisation of language functions depends upon the interaction among the semantic, the syntactic, and the prosodic components of a given stimulus. The authors present a neural model in which syntactic and semantic processing pertains to temporo-frontal networks of the left hemisphere, whereas phrasal prosody processing is subserved by right temporo-frontal areas. In this model, intonation processing, more interestingly, displays a more complex picture; isolated pitch is thought to engage only right hemisphere networks but the more linguistic a stimulus becomes, the more dynamic the involvement of the left hemisphere becomes. Although isolating every single component on which one wishes to focus on might be a valid strategy, isolating prosody from its linguistic context might not facilitate the understanding of the dynamic interaction of different linguistic levels. As was noted by researchers in Ross et al. (1981), Friederici and Alter (2004) also suggest that the interplay between syntactic and prosodic processing can be explained by an interaction between the two hemispheres dependent on the corpus callosum, the structure that connects the two cerebral hemispheres.

The corpus callosum is a complex structure containing fibre tracts that are topographically arranged in terms of their cortical origin (Sammler et al., 2010). This structure seems to be of paramount importance when it comes to prosodic processing. Hence, not only damage to the cortex but also damage to the interconnecting structures can lead to prosodic impairments. Klouda et al. (1988) demonstrated that affective and linguistic aspects of prosody were compromised in a patient with damage in the corpus callosum. These findings stress the integration of information by the two hemispheres as opposed to processing in each hemisphere separately.

More recent findings point to different roles for specific parts of the corpus callosum. As mentioned above, the corpus callosum is structured in such as a way that different fibres correspond to different parts of the cortex. This entails that damage to different parts of the corpus callosum will result in different processing dysfunctions. In Friederici et al. (2007), speech comprehension tasks with patients with different lesions in the corpus callosum showed that the posterior third of the corpus callosum is the most significant anatomic structure responsible for the interaction between processing of prosodic information and syntactic structures.

A cross-linguistic study by Gandour et al. (2004) provides some illuminating findings on the contribution of the two hemispheres and the corpus callosum, highlighting how the lexical component of speech affects the processing of its prosodic realisation. In this study, researchers compared Chinese speakers and American English speakers with no knowledge of Chinese, examining the perception of Chinese linguistic prosody by the two groups. According to their imaging findings, the English speakers showed greater right-sided activity when exposed to Chinese utterances. By contrast, Chinese speakers displayed a larger left-hemisphere

activation when presented with the same Chinese utterances. Hence, being exposed to exactly the same stimulus seems to entail different prosodic lateralisation depending on whether linguistic processing takes place at the same time or not. This type of cross-linguistic experiment allows, according to the authors, the detection of areas that are more sensitive to linguistic processing versus areas more sensitive to processing of lower-level acoustic features.

The study of emotional prosody also pertains to the intricacies of lateralisation of speech processing. Relevant questions have been studied through lesion studies. In contrast to left hemisphere lesions that lead to aphasia, right hemisphere lesions can cause prosodic impairments that appear to be the result of difficulties in matching the emotional and compositional components of language due to intrahemispheric and transcallosal interactions and are referred to as 'aprosodias' (Ross, 1981).

In order to explore the neural substrates of production and perception of emotional prosody Ross et al. (1997) designed an Aprosodia Battery that they distributed to both left hemisphere and right hemisphere-damaged patients. The stimuli assessing perception of emotional prosody encompassed three different sets of tasks with varying linguistic content. The first set included different realisations of an utterance in 5 emotional tones and a neutral tone. Different words were intoned across the stimuli. The second set was similarly developed with monosyllabic sentences ("ba ba ba ba ba ba") and the third set included an asyllabic sentence ("aaaaahhhhh"). This variation in verbal-articulatory content aimed to reveal different lateralisation across participants. Results showed that the performance of left hemisphere-damaged patients improved on the second and third set, whereas performance of right hemisphere-damaged participants remained almost the same across the three sets. Therefore, reducing the purely linguistic demands of the tasks had a facilitating

effect for the left hemisphere-damaged patients and not those with right hemisphere lesions. The authors argue that this was due to the fact that reducing the verbal-articulatory load in the second and third set shifted the tasks from being processed bilaterally to being mainly lateralised to one hemisphere.

The study of the aprosodias has led to the assumption that there is a 'nonverbal affect lexicon' in the right-hemisphere, encompassing emotional prosody, facial expressions, and gestures (Bowers et al, 1993). This type of emotional lexicon, according to the authors, mediates the reading of the social displays among humans, and the affective disorders that have been reported in the literature relevant to prosody, facial expressions, or gestures can be thought as disruptions of this network. Hence, it seems that emotional meaning can be as systematically organised as semantic processes are in the left hemisphere.

Emotional prosody has been studied in a relatively large number of studies and seems to differ from other aspects of the perception of emotion outside the domain of spoken communication. For example, Heilman et al. (1984) studied both right and left hemisphere-damaged individuals on emotional and non-emotional prosody. The emotional prosody contained declarative sentences in three emotional tones, namely happy, sad, and angry. The non-emotional linguistic prosody included declarative, interrogative, or imperative sentences which were marked by changes in pitch, stress, and phrasing. All patients manifested poorer linguistic prosody compared to controls. However, the right hemisphere-damaged patients had a more impaired comprehension of emotional prosody compared to the left hemisphere-damaged participants. The findings of this study, therefore, suggested that the right hemisphere plays a pivotal role in the processing of emotional prosody but this is not the case for the processing of linguistic prosody.

Blonder et al. (1991) also examined left and right hemisphere-damaged patients and found that right hemisphere-damage does not provoke impairments in the ability to infer an emotion when being exposed to a sentence describing a situation. However, right hemisphere damage and disruptions in areas supporting interhemispheric connections can compromise the ability to encode and perceive emotional tone based on prosodic features (e.g., Ross et al., 1997). This is also the case for facial and gestural expressions; for example, in Borod et al. (1986) right hemisphere patients were found to be impaired at perceiving and expressing facial emotions. It, therefore, appears that the prosodic realisation of emotion dissociates from emotion itself and alternative ways of perceiving it, such as through lexical descriptions.

These sections presented some of the evidence acquired by previous research on the different subdivisions of prosody and the perceptual processing mechanisms that have been shown to relate to them. According to the evidence, linguistic and emotional prosody processing has been shown to dissociate. The role of the two cerebral hemispheres differs depending on a series of factors that have to do with the nature of a stimulus or the methodology of a study. The temporal aspects of speech processing differ in that fast acoustical transitions are processed by different networks than those devoted to long chunks of an utterance. Moreover, the relationship between linguistic context and prosodic realisation has revealed additional aspects taken into consideration in the field, highlighting the effect of the nature of the tasks used in the studies. More interestingly, factors such as being exposed to a language that one does not speak can result in different lateralisation involving different patterns of brain activation. In general, the interplay between lexico-semantic and purely prosodic processing has been attributed to the corpus callosum, a structure that connects the two hemispheres and is associated with an

integration of processing. The next section presents the concept of 'expressiveness', revisiting prosodic labels that have been explored in the existing literature.

2.3 Defining 'expressiveness'

This section introduces 'expressiveness', a new variable of suprasegmental analysis that corresponds to a number of acoustic cues that, as it is argued here, contribute to a similar effect in speech and music. These features appear to have a dynamic gestalt-like quality rather than being a combination of isolable acoustic components. It is also suggested that this quality accounts for the variability in speech and music performance, contributing to appreciation of aesthetic value. Hence, no specific mapping of acoustic cues and terms has been attempted here. The approach that is taken in defining 'expressiveness' resembles a 'proof by contradiction' argument. That is, due to its less tangible nature, instead of defining it by describing its constituent parts, this section refers to the prosodic aspects represented in the existing literature that do not capture this additional dimension. In other words, a sense of what is meant by 'expressiveness' is achieved by identifying prosodic features that do not seem to pertain to it.

'Expressiveness' refers to these features, that despite the difficulty in describing them, differentiate an utterance or a melody from counterpart utterances and melodies that evoke less rich aesthetic appreciation. Acting performance and music soloist performance possibly constitute the more representative examples of this acoustic feature as they are interwoven with the production of a highly aesthetic result. This theoretical argument has been supported by the experiments of this study which, as later sections describe, investigated perception of this feature. More

specifically, they explored whether the healthy and the compromised cognitive system can process 'expressiveness' with similar success in speech and music acoustic streams, along with other aspects of prosody across domains.

In order to better appreciate the differences of this proposed prosodic dimension, it is worth looking at other aspects of prosody that are well-established. The following part of the section provides a consideration of linguistic prosody, emotional prosody, and pragmatic prosody in that respect. Building blocks of prosody (pitch, loudness etc.) have been described earlier in this chapter, but these elements are here examined in the context of these three prosodic labels. Some specific examples are also given later in this section in order to show: why the existing prosodic categories in the literature do not provide a sufficient description of all aspects of prosody; why they do not have a one-to-one correspondence to the prosodic realisation of speech; and, why the description of a new prosodic type is, in fact, necessary. As music has not been prosodically described to the extent that speech prosody has, music prosody is here described based on soloist performance features.

2.3.1 Linguistic prosody

In 2.1.3 pitch was treated as an important suprasegmental feature. At the prosodic level, pitch variations can differentiate an utterance spoken in a monotone from another signalling, for example, a question or a request. Pitch carries a variety of functions that provide linguistic information but also speaker information. More specifically, pitch in combination with other features can contribute to the distinction between a question versus a statement and, at the same time, it provides clues to speaker identity such as the sex and the age of the speaker, and possibly elements of

their attitude and their personality (Clark et al., 2000). Also, manipulations of pitch can be used to emphasise a significant piece of information in an acoustic stream (Ogden, 2009). As pitch has particular value for the study of speech, understanding of its role is vital for the appreciation of intonation and prosody. In the case of linguistic prosody, combinations of several acoustic features are employed to mark different linguistic structures. They can be also used to place emphasis on a specific word or part of an utterance.

Although the terms linguistic prosody and intonation are often used interchangeably in the literature (Botinis et al., 2001), there is an important difference between these two labels; intonation only refers to the pitch patterns of an utterance, while prosody is a much broader term. Intonation refers to the use of tonal features bearing non-linguistic meaning, that is, it often reflects the attitude of the speaker or the nature of the message but it does not carry semantic information, as in the case of duration at the segmental level (Lehiste, 1970). Duration and loudness, which have been also described in previous sections, can be appreciated in the broader context of linguistic prosody. Acoustically, similarly to pitch manipulations, additional loudness or longer duration can make a portion of an utterance more emphasised, or 'prominent' (Skandera and Burleigh, 2005). All these features provide linguistically structured 'postlexical' meaning (Ladd, 2008). That is, they do not bear semantic information themselves but do contribute to the communication of meaning.

When one listens to a language they know, they are normally able to perceive individual words in the acoustic signal rather than an undifferentiated sound stream. Some words appear to be grouped with other adjacent words, forming 'prosodic phrases' (Wightman and Ostendorf, 1994). These phrases are not constructed arbitrarily. A way of communication of meaning through linguistic prosody often

emerges in the form of syntactic disambiguation. In other words, as a group of words can give rise to more than one semantic interpretation, linguistic prosodic features can differentiate one version from another. This happens through syntactic segmentation of continuous speech into different syntactic units (Vaissière, 2008). That is, when the lexical content of an utterance does not suffice for successful communication due to multiple possible interpretations, its prosodic realisation plays a crucial role. One form of phrasing rather than another can point to the correct semantic interpretation through syntactic disambiguation. For example, in the sentence "He will cook sausages or fish and fries", the written form cannot inform the reader if the person cooking may make sausages only, sausages and fries, or fish and fries. Grouping "sausages or fish" rather than "fish and fries" together in the spoken version of this sentence would disambiguate the ambiguous content of the written form.

Placing emphasis through the use of variation in pitch, loudness or duration on a part of a phrase also pertains to linguistic prosody and seems to have a counterpart in music. Both in speech and music, some prosodic means are used to highlight the importance of a structure in relationship to other structures of the same context. Prominence of acoustic events, thus, emerges as an expressive contrast among parts of a melody. Acoustic contrasts can also be achieved through variation in articulation (achieved through touch in the playing of a musical instrument) and also through highlighting subdivisions of a phrase (Sloboda, 1983). This tendency resembles the strategies used to achieve contrastive focus in speech in order to favour one interpretation over another (Palmer and Hutchins, 2006). In speech, contrastive focus is often realised through the use of different pitch patterns. These might refer to a rise or a fall in pitch, but also to whether a pitch movement lasts for more than one

syllable (Hermes and Rump, 1994). Contrastive focus, which can be achieved in one of the above ways, depends on what the speaker assumes about their interlocutor's beliefs and expectations and it, therefore, bears an amount of subjectivity (Zimmerman, 2007).

2.3.2 Emotional prosody

As linguistic prosody, emotional prosody also pertains to the melodic and rhythmic speech features that contribute to the listener's understanding of the speaker's disposition (Pell, 2006). Acoustic features can be shared among different prosodic types. For example, loudness or pitch constitutes features of both linguistic and emotional prosody, but their prosodic labels are much more than a mere group of acoustic features. Emotional prosody constitutes the use of acoustic features in order to convey a specific emotional tone, for instance, happy, sad, angry, or fearful. In other words, it is indicative of the speaker's mood (Vaissière, 2008). This type of prosody carries communicative meaning, which is distinct from linguistic prosody. In contrast to referential meaning, which is communicated through the lexical aspect of speech, variation of intonation or loudness can contribute to the communication of affective or emotive meaning (Crystal, 2008).

Hence, everyday communication is informatively enriched by the use of emotional prosody (Nygaard, 2005). However, in contrast to referential meaning that appears to be relatively stable, typically without ambiguity in its interpretation, meaning carried by emotional prosody often appears open to interpretation by each listener. In an experimental context, emotional speech recognition has been tested in a limited way, as most studies focus on five or six basic emotions, which is not representative of the

full range of emotional prosody which a speaker may express (Ververidis and Kotropoulos, 2006). Therefore, even if a study points to consensus in the judgement of some prosodic realisations of emotion, one should take into account that this is a limited investigation of emotional prosody perception, filtered through the experimental rather than a naturalistic context.

2.3.3 Pragmatic prosody

Prosody also encompasses pragmatic aspects of speech, that is, acoustic realisations that depend on the context in which speaking takes place, including the individuals that participate in a given conversation. These features do not necessarily correspond to specific pitch contours as, for example, in the case of interrogative sentences. One example of pragmatic prosody is sarcasm. Sarcasm and other pragmatic features are treated separately from emotion, as, although expression of sarcasm might overlap with the expression of an emotion, such as disgust, it is not itself an emotion. A wide range of other emotions, such as anger or surprise, can also accompany sarcastic output.

Pragmatic prosody might relate to specific pitch realisations but this is not always the case. That is, some pitch contours accompanying the speaker's output are likely to be associated with a specific attitude and not another. For instance, making a request in a polite way has been attributed to specific contours of rising or falling pitch (Wichmann, 2002). However, it can be argued that attempting to create an exhaustive list of prosodic labels for all possible kinds of implied pragmatic meanings would be a difficult project.

In fact, a rather complex picture seems to emerge when it comes to pragmatic prosody. Whenever sarcastic tones, for example, are employed these can vary largely from utterance to utterance. Attardo et al. (2003) explain that both a large and a very small pitch movement can lead to an ironic prosodic expression. According to them, it is not a specific pitch contour that makes an utterance ironic but rather the intonational contour being in contrast to the expected intonation that would normally correspond to this utterance. What is more, the context of pragmatic prosody is often hard to explain and describe, and it is due to the very nature of the context that pragmatic meaning has various prosodic realisations. That is, the pragmatic force of some instances of speech can be so significant that an utterance will be considered ironic even if its prosodic colouring is completely neutral (Cutler, 1974).

These arguments highlight an important difference of pragmatic prosody from linguistic and emotional prosody. In the case of pragmatic prosody the use of suprasegmental cues is much more flexible and often unnecessary. Employing a neutral prosodic tone to signal a question would be almost impossible, except if the question is signalled lexically, introducing the listener to the fact that a question will follow. In contrast, pragmatic prosody expressing irony can be conveyed flexibly; using a prosodic colouring opposite to the accompanying lexical context or even omitting any kind of prosodic colouring. Hence, this type of prosody appears to rely heavily on its context.

2.3.4 Beyond the existing prosodic labels

The known prosodic labels correspond to some conventions in speech but also in music that justify their systematic description by researchers. For example, it is

undeniable that prosodic variation in both language and music serve as building blocks of a given structure. According to Grosjean et al. (1979), the constituent phrase structure of a sentence might work as a predictor variable accounting for variation in the allocation of pauses. They created a model that could, in fact, predict pausing based on syntactic structure. Similar models have been designed for musical prosodic features. Todd (1985), for example, offered empirical support for a similar argument in music. His study suggested that the degree of slowing of tempo at the end of a phrase is the result of the performer's reflection upon the hierarchical structure of the phrase. Hence, as in phonological units of speech, slowing down of the tempo in music towards the end of a musical phrase and often lengthening of accented tones is present in musical units (Repp, 1992). It seems that structure does, indeed, relate to the determination of pausing and rate across domains. In other words, the use of prosodic cues is conditioned by phrase structure and it is, therefore, meaningful to describe, for example, linguistic prosody in order to better appreciate the structure of spoken utterances.

It should be appreciated that speech provides richer information compared to written language. In other words, lexico-semantic information, along with conventions of punctuation, does not entirely suffice for a prediction of the acoustic realisation of a sentence. In both speech and music, there can be acoustic manipulations that do not change the categorical information carried by a sentence or a melody, that is, words and musical pitches respectively (Palmer and Hutchins, 2006). Although this study is not about production but perception, references to how sentences are likely to be produced based on written information will be crucial in order to highlight the fact that additional information other than linguistic and emotional prosody is included in the speech signal.

An example from speech prosody illustrates why syntax cannot fully predict the prosodic realisation of a sentence. That is, not being able to predict the prosodic realisation of a sentence based on the syntactic information provided suggests that other elements beyond syntax might account for the suprasegmental organisation of a spoken utterance. In Brown and Miron (1971) a professional speaker was asked to read a passage. Results showed that only 64% of pause duration variance could be predicted from the syntactic structure of the sentences constituting the passage. Another example comes from length relationships in utterances. Ferreira (1991) also showed that syntactic structure cannot serve as a sufficient predictor of word durations and following pauses. Her research suggests that syntactic structures can relate to timing patterns but this is not a transparent relationship, which means that syntactic constituents do not correspond to duration counterparts in a one-to-one fashion.

These examples illustrate that, despite the effect that syntax can have on some aspects of the acoustic realization of an utterance, the sentence does not provide all the necessary information in order to fully prescribe prosodic performance. It, therefore, appears that there should be additional features that account for the unpredictability in a speaker's output. This is even more likely a consideration in attempting to describe the variation informal or professional speech performance that is likely to have higher aesthetic value.

In a similar vein to written scripts and a rendition of its spoken performance, there is little predictability between music scores and a given musical performance. According to Palmer and Hutchins (2006), it is not obvious if music does have a

strict grammar, as there are some aspects of expression in performance that do not seem to reflect known principles. The authors argue that studying musical structure is separate from studying music performance which, in our case, falls under the scope of 'expressiveness'. As spoken utterances display features that are not present in their written counterparts (Shattuck-Hufnagel and Turk, 1996) and written musical notes differ significantly from their acoustic realisation in a musical phrase, it can be suggested that the acoustic cues that are present in the speech and the music signal appear to go further beyond a syntax-prosody relationship. One could not, therefore, predict an actor's rendition when they read a script based purely on the text.

With respect to the constraints governing rhythmic patterns, these do not seem to apply equally across domains, as was discussed in 1.2.2. In music, there might be an unvarying occurrence of a set of beats that get repeated. Although strong and weak beats appear across both language and music, the degree the two domains stick to this pattern differs. Speakers do not seem to produce an output that can be strictly mapped into a repetitive strong-weak temporal pattern.

There have been arguments in favour of isochronous speech referring to different language categories according to their rhythmic organisation, namely syllable-timed and stress-timed languages. This distinction, however, has been shown to be of limited utility and only a small reference will be made here, as the focus is on showing why prosodic conventions do not have a systematic application rather than delving into this typology. In order to better understand the distinction between syllable-timed and stress-timed languages, it is important to give an account of rhythm. In music, rhythm is defined as the regular organisation of temporal relationships in a musical measure that produces the perception of strong and weak

beats (Zatorre et al., 2007). In speech, rhythm constitutes the recurrence of prominent stresses or syllables that the listener perceives at regular intervallic moments, and in the case of some oriental languages, high pitches (Skandera and Burleigh, 2005). Based on its 'rhythm', a language can be considered as 'stress-timed', meaning that there are syllables with longer duration that are more prominent than other syllables, and these prominent syllables appear to occur at relatively equal intervals (Clark et al., 2000). Hence, this definition implies that there is a regular recurrence of stressed syllables in this group of languages. In contrast, some languages are considered to be 'syllable-timed' where there are no durational contrasts and the total duration of a phrase depends on the number of the syllables it contains. Roach (1982) provides a helpful account of some interesting reasons why this classical typology is not accurate and reliable. He argues that there is substantial disagreement among phoneticians about syllabification, stress, and intonational units even when they study their native language. One should not confuse, he continues, the subjective impression of language rhythm with objective instrumental measures. Hence, one cannot create a typology based on the perception of the listener, despite impressions of 'isochrony' in speech, as they do not, in reality, resemble the beat patterns which exist in the concrete measures in music.

It is meaningful to turn, at this point, to the proposed feature of 'expressiveness' in order to examine if this prosodic feature can account for the prosodic elements that are not fully mapped into the existing prosodic descriptions. Although two actors would use similar intonational cues to signal, for instance, a question as opposed to a statement, there are various cases where one could not expect two actors to produce an identical interpretation of a sentence or a passage. Likewise, in music, several expressive guidelines are provided in a musical score regarding tempo, loudness and

often emotion, but these categories are not exhaustive, as there is always room for variation (Palmer and Hutchins, 2006).

One can argue that dramatic actors employ a different range and degree of prosodic cues when performing as compared to casual conversations in everyday life. Artistic performance might call for different phrasing patterns and different variations in length and loudness. This does not mean that prosodic rules are violated but rather that the performer chooses to employ, for instance, longer vowel duration patterns then would be typically employed in casual speech in order to make their output more interesting acoustically. Manipulating acoustic signals to achieve an expressive result by musicians resembles the speech manipulations by talkers (Palmer and Hutchins, 2006). Extending this argument, one could suggest that musical prosody resembles more artistic speech prosody (an actor's output), as in art, emphasis of prosodic patterns might account for its aesthetic value.

In fact, performers do not achieve an expressive interpretation by using identical prosodic patterns. Likewise, in speech, and more specifically in the vocal delivery of actors, despite a number of conventions, different actors will use different vocal effects to deliver the same passage, and even the performance of the same passage by the same actor will differ from time to time. Despite the need of appropriate segmentation, emphasis, or creation of emotional states, a listener might appreciate more a given performance compared to another and consider an actor more talented than another due to other such intangible cues. The existence of alternative prosodic manipulations might suggest the existence of a prosodic dimension beyond syntactic relationships. Some of these features are linked to syntactical hierarchies. However,

some of them are not necessarily structure-dependent or they can, alternatively, be interpreted from an aesthetic point of view.

In music, deviation from temporal regularity can be achieved through a number of techniques such as 'rubato', 'accelerando', or 'rallentado' which are employed in different ways, as more experienced players are likely to make a different use of prosodic features in their performance (Sloboda, 1983). Rubato more generally pertains to the performer's choice of not following the notated durations by both increasing and decreasing duration according to the events that should be emphasised. 'Accelerando' more specifically refers to the increase in tempo in a certain part of the phrase and 'rallentado' as the decrease of tempo. Although these labels suggest a temporal tendency of a specific type, they allow for variation among performers. In their definition of the functions of musical prosody, Palmer and Hutchins (2006) refer to the segmentation of an acoustic signal into constituent parts, the emphasis on the most important events, and the creation of emotional meaning. At the same time, they suggest that there are no deterministic rules when it comes to using some prosodic cues and performers can be flexible in terms of the use of these cues or the degrees to which they use them, meaning that prosodic manipulations might not be judged as incorrect but rather as less preferred. One could, therefore, suggest that there is an additional important function of musical prosody; that of aesthetic appreciation. In the speech domain, although terms such as 'rubato' or 'accelerando' do not officially prescribe an actor's performance, similar manipulations occur in acting that can lead to a more or less preferred aesthetic result.

One could define a performance in speech and music as 'expressive' as the performance in which a combination of prosodic features differentiates it from a deadpan version of the same content. The deadpan version can be thought of as a speech or a musical phrase with the minimal amount of prosodic cues resulting in an aesthetically 'neutral' result. Also, a deadpan version would have little variation of these features, having, for example, evenness in loudness, pause placement, and pitch (only in the case of speech) patterns across a phrasal chunk. In contrast to syntax or to the intention of conveying a particular emotion, 'expressiveness' appears more related to an optional preference rather than some absolute judgement of correctness. This seems to be a distinctive feature of 'expressiveness' in comparison to structural (linguistic in the case of language) and emotional prosody. It therefore, seems that prosodic structure does not necessarily depend on syntactic structure and can consequently stand on its own, being independent of grammar rules. That is, although performance both in the speech and the musical domain falls within some restrictions (e.g., one would not, for example, use a declarative statement contour when asking a question), a large amount of variation occurs among performers. Hence, speech and music are not exclusively characterised by purely deterministic rules, thus allowing for variation among individuals.

Apart from the limited role of structure, it is worth taking into account the fact that emotional prosody in speech and music does not provide a complete account of the prosodic richness of speech and music acoustic streams, thus supporting the claim that this additional dimension of 'expressiveness' is needed for a fuller description. If research has demonstrated little consensus on specific emotions conveyed to the listener by a particular utterance, then attempting to label expression according to

such basic emotions might not be a meaningful process. It is undeniable that emotions relate to various forms of expression. However, the question arises as to whether there are cues that make an utterance or a melody emotion-specific or not. To put it a different way, it is worth investigating if aesthetic expression always overlaps with the expression of a specific emotion or if it can be achieved independent of a given emotional tone. In the case of emotional prosody, a listener is able to processes emotions even when the lexico-semantic content of a given utterance has been removed (Scherer et al., 1984). In a number of studies, emotional and linguistic prosody have been shown to dissociate. What is more important, however, for the purposes of this study is to consider additional dimensions that lead to appreciation of an acoustic output. Is it only structure and emotion that account for the appreciation of speech and music? Does aesthetic appreciation rely on structure and emotion or does it also rely on 'expressiveness' beyond emotion and separate from structural frames?

If prosodic 'expressiveness' is in some ways unrelated to the formation of specific and invariable syntactic structures, and if it does not wholly pertain to the expression of emotions, it seems that both speech and music prosody display an additional dimension which has not received sufficient attention. This further dimension might be harder to grasp and to define as it does not seem to be accompanied by specific labels of structure, meaning types, or emotion categories. However, this does not suggest that it should be overlooked. According to the above arguments, prosodic output can be predicted on the basis of syntax only to a certain degree. Research has shown that prediction based solely on morphosyntactic structures is insufficient. Speech and music prosody both seem to embody this property. Constraints have

been described in both domains but they do not provide a comprehensive account of the observed variation in expression. Moreover, the existent prosodic labels tend to overlook the fact that acoustic cues interact and, therefore, should not be treated in isolation. Loutrari and Lorch (in press) argue that while breaking down prosodic streams into identifiable components has been of critical importance for the study of speech and music prosody, this type of analysis excludes from consideration dynamic aspects of speech and music.

The concept of 'expressiveness' has been defined here through the consideration of the acoustic features that it does not encompass, that is, describing features that have been associated with syntax or emotional prosody. As prosodic manipulations do not seem to solely play the role of syntactic or emotional markers, 'expressiveness' has been suggested to constitute a set of acoustic features that add to the aesthetic result of prosodic modification. If prosody is defined as the set of features adding to emphasis of the significant acoustic events or the trigger for emotions, 'expressive' prosody is the acoustic constituent that adds to aesthetic appreciation. Research with music performers has shown that despite the general principles they might follow and the set of restrictions that they cannot break, they often manipulate the prosodic output of their performance in such a way that their performance becomes unique. It was argued that the written and acoustic forms of language and music do not have a transparent correspondence which links to the previous argument on variation among performers. An analogy in prosodic choices has been shown to exist in speech, as several experiments reviewed here have provided evidence that a reader cannot predict the prosodic choices of a speaker when it comes, for example, to pausing or length manipulations.

The above arguments highlight the significance of studying these additional aspects of prosody that are not captured in the existing literature. It appears essential to explore these additional acoustic cues that do not seem to be predicted by the description of syntax or emotion. Additional work with healthy participants and individuals with compromised cognitive performance will be critical in the novel investigations presented here in order to detect and define different pathways that might be involved in the processing of 'expressive' prosody.

2.4 A note on case study methodology: theory and contribution to the present study

As the present investigation involves two case studies, the theoretical underpinnings of the case study methodology will be considered. These include the advantages of the specific methodology, the criticisms that have been raised and the limitations that are inherent in it. The way that researchers use this methodology in order to study cognitive impairments is presented, as well alternative interpretations that might account for their behavioural results obtained by the study of brain-damaged individuals. These methods will be discussed in the neuropsychological framework that has been proposed by leading figures and developed over many decades. At the theoretical level of consideration, the reasons that account for the choice of this methodology are discussed in 2.5. Moreover, the effect that the nature of tasks used across studies can have on the obtained results is analysed in 2.6.

Case study methodology has been extensively employed in neurolinguistics and neuropsychology. Its contribution to these fields has been both praised and criticised. Case study methodology is here considered along with the inherent complexity of the interpretation of the findings arising from it, in an attempt to evaluate its role in

the comparative framework of language and music processing. Consideration of both the hurdles and the advantages of this approach are presented.

2.5 Single case study design

In the exploration of the function of the cognitive system, case studies compare the performance of brain-damaged patients to that of neurotypical individuals on a series of tasks. In order to infer the presence of an impairment in a patient due to brain damage, case study researchers typically form a hypothesis prior to testing and if a substantial difference appears between the patient and the control group on a given task, this suggests that the patient displays a deficit on the task in question (Crawford and Garthwaite, 2007).

This section mainly focuses on the significance of case studies with brain-damaged individuals, in order to show the significance of focusing on the performance of the brain-damaged individual of this study. As the systematic study of congenital amusia is more recent compared to the investigation of acquired disorders, the treatment of congenital amusia methodology in terms of individual differences is here limited.

Studying individual cases of brain-damaged patients often reveals a complex picture of patterns of deficits that co-occur with preserved abilities. That is, the presence of a deficit does not rule out the possibility of performance within normal range on a different task. Hence, a 'dissociation' occurs when a patient's performance on a task is extremely low compared to a group of control participants and, at the same time, their performance is at a normal or at least better level on another task (Shallice, 1988). A dissociation is also observed in cases where performance is impaired on two given tasks, if it is significantly more impaired on the first task compared to the second task and vice versa (Coltheart, 2001). Given the complexity of the cognitive

system, it seems that studying a large number of single cases can contribute to the exploration of numerous combinations of intact and compromised abilities.

To better appreciate the value of the single case study approach, its contribution should be considered in the appropriate framework of exploration of human cognition. Caramazza (1986) argues that the purpose of the single case study approach is to generalise to the normal (used here in the sense of typicality) cognitive system and not to a particular patient population. If one assumes that there is a model of 'normal' processing, the observed patterns in a brain-damaged individual can contribute to the assessment of a given model. It is, however, important to define the context in which this argument appears to be valid. Coltheart (2001) suggests that conclusions from single case study impairments are safe if one assumes that the architecture of cognition is constant across all healthy individuals. To further illustrate this point, it is worth referring to the 'assumption of universality' (Caramazza, 1986). According to this assumption, the healthy brain performs in a way which is true of any individual brain. Thus, the presence of brain damage provides a way to compare the 'universally' healthy brain with that of an individual suffering from brain damage. In general, in order to make inferences from pathological to normal language, it is assumed that the performance of brain-damaged individuals depends on mechanisms that are preserved and not mechanisms that are developed as a result of brain damage, although it is not possible to completely rule out this possibility (Saffran, 1982).

It can be argued that the need for conducting single case studies can be more clearly understood when one considers the inherent problems in the alternative of using group studies. As linguistic and in general cognitive disorders can be broken down to impairments of smaller processing units, it appears impossible to examine them in

detail by studying groups of individuals that are homogeneous on all relevant variables. In fact, speech disorders do not always occur with identical symptoms across different patients. Analysing smaller components can be particularly informative, as such disorders may be quite selective and it is common to find some language functions to be disrupted and, at the same time, others to be intact. As associations and dissociations of impairments vary across patients, obtaining a detailed picture of a specific impairment through group studies might prove problematic. For example, it might be easy to find a group of people diagnosed with Broca's aphasia but it is unlikely to find a group of patients with unimpaired sentence comprehension and sentence construction but impaired ability to generate affixes at the same time (Caramazza and Coltheart, 2006).

It, therefore, seems that the case study approach provides a reliable alternative for the investigation of the cognitive system to that of group studies that typically assign patients to categories. Attempts to classify patients in groups can be misleading as it might be hard to rule out every possibility of sharing features that are traditionally related to other groups. For example, Schwartz (1984) argues that classifying aphasic syndromes with the purpose of generalising to the architecture of language in the brain is not always trustworthy, as findings from a given task can reveal deficits or spared processing which may or may not be related to the category to which patients had been originally assigned. Likewise, Caramazza (1984) refers to the practical problem of studying patients on the basis of classification. He suggests that group studies typically involve classifying patients in the same group without them displaying exactly the same deficits due to the lack of sufficiently detailed inclusion criteria. Hence, attempts to look at homogeneous groups or to classify patients according to a fixed set of characteristics seem to overlook the unique properties

which can be related to a given brain lesion. That is, testing the performance of patients in a group does not provide an insight into the particularities of every individual lesion site. By contrast, detailed assessments of individual cases allow for the consideration of unique patterns of associations and dissociations from each natural experiment caused by the particular pattern of pathology.

As group studies and single case studies tend to have different objectives, their tools can vary along various stages of research. Replication seems to be one of these tools. While the findings coming from a single case study are supported by a smaller amount of data compared to large scale studies, Caramazza (1986) argues that replication, typically valued in group study research, is only necessary if one accepts that the conditions of the initial experiment are unclear or encompass many uncontrolled factors. The issue of replication can be also examined in relation to the structure of the cognitive system. Coltheart (2001) points out that replication is not necessary to validate a processing pathway found in a given patient if the function of the cognitive system is accepted to be uniform. In this framework, if a patient is found to display a given dissociation, the inability to replicate this processing dissociation in other patients does not necessarily disprove the initial findings.

The above points relate to criticism found in the literature. For instance, although Schwartz and Dell (2010) acknowledge the contribution of single case studies to neuropsychology, they argue that the single-case approach does not give the opportunity to analyse patients as a group and detect quantitative trends in a sample providing the opportunity of making predictions about these trends. According to Caramazza's (1986) argument, the criticism that one cannot generalise from case studies to a patient population is not valid, as generalising to a patient population is out of the scope of the single case study approach and its real purpose is to generalise

to the neurotypical cognitive system. One can argue that predicting the presence of a particular impairment in a wider population does not necessarily constitute the most appropriate way to explore the structure and function of the healthy brain. In general, it appears that findings should be examined in a framework specific to a given methodology in order for them to be interpreted and appreciated accordingly.

There are some more meaningful arguments pertaining to the findings emerging from case studies which might render interpretation of dissociations less clear and straightforward. Chater (2003) argues that dissociation in the performance of two tasks does not necessarily indicate that the two tasks are related to two independent cognitive mechanisms, and might alternatively suggest that tiny specialised components might account for differences in performance. He also argues that different tasks might require the participation of a processing component in varying degrees, thus resulting in perhaps sufficient performance on a task and insufficient performance on another. Therefore, in this case, a dissociation in performance does not necessarily entail the existence of two different processing components, but rather a varying degree of 'specialisation' of the same processing component.

Alternatively, the better performance of a patient on one task compared to another might simply suggest that one of the tasks is much simpler than the other (Shallice, 1988). For example, a patient might be impaired at understanding printed words but able to understand spoken words and this can be interpreted either by arguing that there are two distinct pathways related to these processes or that there is a single module serving word comprehension in general and comprehension of written words might simply be more demanding compared to spoken word comprehension (Coltheart, 2001). It can be suggested that interpreting the findings of case study research in the light of additional sources of related evidence can provide the

opportunity for an in-depth analysis of the results arising from case studies avoiding, at the same time, simplistic interpretations.

Case study methodology meets the needs of the present investigation for two main reasons; namely the originality of the acoustic patterns that were studied and the unique profile of the brain-damaged individual and the congenital amusic that participated in this study. The latter reason relates to the fact that the brain-damaged patient displayed an extensive unilateral right hemisphere lesion sparing the right frontal cortex (shown to be involved in the retention and comparison of acoustic patterns) and to the fact that he had some pre-stroke music experience. With respect to the inclusion of the congenital amusic individual, it should be noted that, despite the general diagnosis of amusia, the performance of congenital amusic individuals on cognitive tasks does not necessarily follow predictable patterns.

It can be argued that, as several decades ago, the aphasia labels that were used resulted in overlooking particular aspects of individual cases, a similar trend might exist in the present days in studies of congenital amusia. Hence, if a given individual scores poorly on a battery of evaluation of amusia, they are automatically diagnosed as 'amusic', without being compared to other individuals on specific musical processes. If the plethora of studies on congenital amusia recruit groups of congenital amusic participants, individual differences will be harder to interpret. As discussed in 1.6, variation in the performance of congenital amusics was detected in Patel et al. (2008). This can be either due to methodology or actual differences among congenital amusics. In both cases, these differences call for individual testing of congenital amusic individuals on additional tasks apart from their mere evaluation as amusics or non-amusics.

The combination of acoustic features investigated through the present tasks constitutes an original contribution to the parallel investigation of language and music. No previous research related to the features explored here has been identified and, therefore, no links have been previously established regarding such variables. Hence, the single case study approach was deemed as an essential starting point for the investigation of an additional dimension of prosody across language and music and the identification of shared or distinct processing across domains.

Summarising the above arguments, it seems that single case study research can provide an enlightening insight into the structure and function of the human cognitive system, as long as the interpretation of its findings is made according to an appropriate framework, without, for example, claiming to predict deficits in a wider patient population. While case study research does not seem to allow strong generalization of findings to make predictions for patient populations with similar characteristics, its value relates to a careful analysis that does not overlook the uniqueness of a given brain lesion or a neurodevelopmental deficit. The case study approach provides the opportunity to investigate impaired performance and brain lesions sites, testing single cases against a model. If one assumes that the human cognitive system can be analysed into smaller processing units, studying the performance of neurologically impaired individuals in homogeneous groups appears impossible. By contrast, looking at selective impairments seems to be a more powerful method to obtain detailed results without failing to attend to the complexity of the human brain.

2.6 The effect of task design on cognitive processing

As a new set of tasks was designed for the purposes of this study, it is important to appreciate the fact that results on performance can often stem from the particular nature of a given task. Given that some of the present tasks are novel, no direct comparisons with previous literature can be made. As far as the rest of the tasks are concerned, comparisons between the present study and previous studies are –to some degree– restricted by the assumption that methodological variables might have an influence on the outcomes. This section, therefore, gives an overview of methodological variables that have been found to affect findings in a number of studies.

Understanding methodological effects entails the appreciation of the specific context a question is put. Therefore, comprehending the type of processing of a given acoustic feature will be interwoven with its experimental context. For example, perception of linguistic pitch and musical pitch, comparison of different pitch sequences, and perception of missing frequencies appear to be three distinct processes and the questions arises as to whether they relate to different cognitive mechanisms. Or, pitch tasks of the same kind can be less or more complex in terms of working memory demands and this is also likely to affect performance. It logically follows that tasks examining a different aspect of pitch perception cannot be necessarily used as evidence for all different types of processing related to pitch.

The following two examples show that a more complex parameter in an experiment does not simply make a task more difficult but might also engage different or additional brain regions. According to Warrier and Zatorre (2004), when one is required to perform a simple pitch task, the left temporal lobe is significantly

involved, whereas on a more complex task, the right temporal lobe has a dominant role, as shown by the study of right and left temporal lobe patients. For example, a task requiring participants to compare two pitch events might be performed by the left hemisphere, but if additional interfering tones are added, the task's complexity calls for the participation of the right hemisphere. This, according to the authors, is due to the role of right hemisphere areas in processing pitch relative to previously presented tones and in subserving working memory through communication with the frontal lobe.

In a similar vein, pitch sequences of the same duration might differ in terms of how pitch patterns are organised. In a functional magnetic resonance imaging (fMRI) study, Hyde et al. (2008) presented healthy participants with pitch sequences which they listened passively. In some of the stimuli, pitch distances were smaller than in others. Results indicated that when a sequence included large pitch changes, the left auditory cortex was involved but when it came to smaller intervals, the left auditory cortex did not get activated and only right hemisphere regions were involved. Hence, processing of pitch sequences varies, depending on whether pitch movements are large or finer. In both of the cases described above, the right hemisphere appears to assume an important role when the processing demands imposed by a pitch task increase.

Direction of pitch and tonal context also seem to affect performance. For instance, perceiving the direction of pitch change was shown to be much more demanding than simple pitch discrimination in Johnsrude et al. (2000). Patients with temporal lobe excisions in Heschl's gyrus had greater difficulty in perceiving the direction of pitch change compared to healthy controls. Despite their poor performance on this task, mere discrimination of pitch appeared to be normal. In another study the task

was varied by presenting a pitch stimulus in a tonal context. This also seems to provoke significant participation of the right auditory cortex. This is shown by Warrier and Zatorre (2004) who examined the effect of tonal context in pitch constancy tasks. Pitch constancy refers to the perception of complex tones as having the same pitch when their fundamental frequency is the same but they have different frequency spectra. The researchers investigated the performance of patients with excisions of the right and left temporal lobe as well as matched controls, attempting to determine whether the presence of tonal context can facilitate pitch constancy judgements. When right temporal lobe patients were presented with pitches in a tonal context, they did not manage to improve their performance; this was in sharp contrast to successful performance of this task by patients with left temporal lobe damage and healthy controls. This finding suggests, according to the authors, that areas in the right auditory cortex make pitch judgements relative to previously presented tones. That is, the role of the right auditory cortex in facilitating pitch processing in the presence of melodic contextual cues appears critical. The importance of this particular task in this study lies in the fact that if researchers had only used a task mode in which pitch constancy was assessed with isolated pitch events, this difference between right hemisphere and left hemisphere patients would not have emerged.

Different aspects of speech would not be efficiently examined if tasks with different parameters were not designed. Laterality tasks have pointed to a dominant role for the left hemisphere on linguistic tasks and also an active role of the right hemisphere on paralinguistic tasks involving, for example, emotional prosody or facial expressions (Gazzaniga et al., 2002). Thus, the degree of activation appears to

depend on the nature of a given task and its demands. The more segmental information is included in the task, the more left hemisphere activation can emerge.

When examining perception of prosody, task design appears crucial. As speech encompasses a large amount of different features that often develop in different time domains (see 2.2. for Poeppel's (2003) 'asymmetric sampling in time' hypothesis), it is hard to disentangle these elements from each other. Often findings of a study on prosody might not be pertinent to prosody *per se* but to prosody in a lexico-semantic context. The presence or absence of segmental information when perception of prosody is examined appears to bring about a critical difference among different stimuli conditions. For instance, Meyer et al. (2004) tested healthy participants in three different conditions of prosodic stimuli. One condition included normal speech in which both segmental and suprasegmental information was provided. Another condition included segmental information but pitch variation was removed and a third one provided participants with suprasegmental cues but segmental information was totally missing. Exposure to normal speech yielded activation in various areas of the left hemisphere and only small right hemisphere activation in the primary auditory cortex. When only the intonation contour of a stimulus was provided, different areas, including subcortical structures were activated. Relative to normal speech, 'flattened speech' (speech deprived of pitch variation) was found to activate areas of the superior temporal gyrus and planum temporale in both hemispheres, areas that did not display such a strong response in the case of normal speech. This example indicates that the mode in which intonation is examined can clearly have an effect on brain activation. Applying these findings to the study of brain-damaged populations is likely to lead to differences in compromised abilities, depending on which mode is chosen.

An interesting example of the importance of task design in the parallel study of speech and music comes from the exploration of the performance of individuals displaying congenital amusia (for more on congenital amusia see 1.6). It is here shown that designing additional tasks in order to investigate why the deficit of congenitally amusic individuals was thought to be specific to music according to early findings provided the key to deeper understanding of the deficit. Patel et al. (1998) designed a series of tasks with which they initially tested brain-damaged patients and not congentially amusic individuals. They designed two groups of stimuli; linguistic and musical. The first condition of the linguistic stimuli included spoken sentences of identical lexical content that were pronounced either as questions or statements. The second condition encompassed linguistic pairs in which utterances differed in terms of which word was accented. The stimuli of the last condition differed in terms of grouping that led to different syntactic structures. The conditions of the musical or 'music-like' stimuli were fewer than the linguistic ones and were generated from the fundamental frequency and timing patterns of their parent linguistic stimuli. In Patel et al.'s (1998) study, the brain-damaged individuals with acquired amusia performed either well or poorly on both. That is, one patient was impaired at both the linguistic and musical tasks and the other performed normally on all tasks. This led the researchers to the conclusion that the linguistic and the musical condition of these stimuli involve the same processing mechanisms.

Investigating perception mechanisms in language and music, other researchers used these tasks with patients that displayed either congenital or acquired amusia (e.g., Ayotte et al. 2002, Nicholson et al. 2003, Patel et al., 2005, Hutchins et al., 2010). In Ayotte et al. (2002), for example, amusic participants were shown to have great difficulty in discriminating among the musical stimuli but their performance was

good when it came to discriminating sentences based on their intonation contours. Peretz and Hyde (2003) attempted to explain the impairment of congenital amusic individuals at perceiving pitch sequences in music in the presence of normal processing of intonation. Their suggestion was that the tone sequences that were generated from the linguistic stimuli were more refined in comparison to the pitch contours of the linguistic stimuli that were coarser.

However, Liu et al. (2012) employed an additional experimental condition that provided a more reasonable explanation. They looked at whether discrete pitch stimuli would be processed differently from gliding pitch stimuli. Their results suggest that individuals displaying congenital amusia had greater difficulty with discrete pitches compared to gliding pitches. This, according to the authors, accounts for the seeming domain-specificity of congenital amusia. That is, as discrete pitches are used in music but not as much in speech, impairment at perceiving discrete pitches will point to a music deficit and preserved intonational processing. However, their study differs from previous investigations in that they chose to use stimuli of a short duration of 250 ms which is far from the time domain in which intonation processing is examined. Hence, this study gave an answer with respect to the nature of pitch sequencing that is hard for amusic individuals, but it is not known whether employing longer stimuli would possibly have a different effect.

What needs careful examination is the nature of comparison that some of the above studies have in fact made between speech and music. As all 'musical' stimuli were generated from 'parent' linguistic stimuli, researchers compared linguistic stimuli to a delexicalised version of speech stimuli instead of true musical stimuli. Section 5.2 discusses this issue in more detail, commenting on whether processing of music-like stimuli can be any relevant to processing of real music stimuli and raising the

question of whether comparisons between the domains of language and music can benefit from stimuli that are strictly manipulated rather than naturally occurring in one's environment.

This section referred to several issues relating to the effect task design can have on studies looking at speech and music or at speech-music relationships. Varying the complexity of a task, even when the same acoustic feature is examined, often points to activation of different brain regions. In other cases, presenting a stimulus isolated or in a context can bring about additional differences in processing. The time domain in which a stimulus unfolds is also of paramount importance in a task and might produce significantly different results, even by changing the lateralisation of a process despite the fact that the same element might be tested. In the case of speech prosody, the presence or absence of lexical information can produce important differences in processing. In general, refining investigations with such additional parameters can give a deeper insight into cognitive processing and often invalidate earlier hypotheses. With respect to the comparative study of language and music, it appears important to assure that authentic speech and music stimuli are employed instead of extensively manipulated versions of naturally occurring stimuli.

2.7 Research questions

The questions presented below mainly relate to the two cases that were studied; the right brain-damaged patient and the congenital amusic individual. However, there was also interest in the performance of neurotypical individuals. That is, healthy participants did not only serve as controls for the two cases but also as controls for the new research tool that was designed. References to 'right hemisphere damage'

and 'congenital amusia' in this section relate to the two single case studies. Although the questions below are formed in a more general way, they do not apply to right hemisphere-damaged and congenital amusic populations but rather to the two cases representative of these populations that were studied.

The most crucial research question of this investigation relates to whether 'expressiveness' is a quality perceived in a similar way across language and music. Through the use of tasks that will be described below, the study addresses the question of whether being expressive in speech is perceived in a similar way to being expressive in music. More specifically, the question arises as to whether 'expressiveness' is perceived in the presence of extensive right hemisphere damage and, more importantly, whether perceiving this quality in the speech domain entails similar performance in the music domain and vice versa.

The same question applies to the context of congenital amusia. That is, the present investigation explores whether failing on the tests of an established and widely used Battery of amusia evaluation (Peretz et al., 2003) entails inability to distinguish among expressive and non-expressive melodies. Moreover, the study examines whether the existing amusia evaluation tools constitute exhaustive methods of music cognition evaluation. A question addressed in the same context is whether ability to identify speech in delexicalised stimuli and ability to discriminate between expressive speech and non-expressive speech stimuli is intact or not in the presence of congenital amusia.

An additional research question pertains to whether perception of pitch differences present over long chunks of utterances are perceived in the presence of right hemisphere damage and congenital amusia. As the bulk of previous research on pitch

perception focuses on manipulations of individual pitch events rather than large pitch chunks this question was further investigated here.

Finally, another question addressed in the study relates to whether the existence of musical experience affects perception of 'expressiveness', emotion, pitch, and rate. This question is also specific to the participants of the study and does not address all levels of music experience.

3. Methodology

This chapter presents the methodology of this study, describing the tasks that were designed and used as well as its participants. More specifically, after introducing the task development and two groups of neurotypical individuals that took part in this study, the cases of two particular individuals are presented. The first is a brain-damaged patient and, therefore, details on his history of illness and neurological impairment are provided. The second case is an individual with a developmental disorder, namely congenital amusia, who was also given the same battery of tasks.

The novel tasks that were designed are presented and individually described. These tasks are discussed in terms of the original contributions to the parallel study of language and music but also in terms of some methodological issues that they raise. Also, the research goals for which they were designed are discussed. Details on the technical manipulations that were applied are also provided. Finally, the experimental procedure is described and the precautions that were taken in order for the patient to feel comfortable and in order for confounding factors, such as fatigue, not to bias the results.

3.1 Initial task development

During an initial stage of task development, 3 of the tasks were piloted with a group of 5 younger individuals (age range: 24-28) in order to trial the stimuli that were designed and the testing procedure in general. These controls were not matched to the cases that were studied here but were engaged in order to facilitate the testing process with the matched controls that were subsequently recruited.

The three tasks that were piloted at this phase, namely the speech prosody detection task, the expressive speech prosody task, and the expressive music prosody task revealed an important effect of the order of presentation to the participants. Results on the speech prosody detection task suggested that participants were primed by the stimuli of the expressive speech prosody task which, was presented to them first. As this was a preliminary piloting stage, participants were encouraged to participate in a debriefing at the end of the session where some of them mentioned that they recognised that the manipulated versions of the expressive speech prosody task were the same as those used in the speech detection task. That is, participants' perception of the prosodic features in a lexicalised form in the expressive speech prosody task facilitated their processing in the speech prosody detection task that was later presented.

These reports by the small group during this piloting phase resulted in a change of the order of presentation of these tasks. The speech prosody detection task became the first task of the battery, the expressive music prosody became the second one, and the expressive speech prosody task became the third task.

3.2 Task development: rationale

The tasks of the Battery designed for this study aimed to bridge a gap in the parallel evaluation of language and music cognition. As discussed in 3.3, previous research used speech stimuli and music-like stimuli derived from speech in an attempt to discover parallels and differences in processing across domains. It was here hypothesised that music-like stimuli cannot necessarily offer a reliable measure of music perception as deriving music stimuli from speech can possibly bias their processing. It was thought that testing participants with real music stimuli would be more meaningful. Thus, although the present stimuli were less comparable across domains, findings from them would result in more valid claims about actual music processing. At the same time, the present Battery would add on the percepts tested through a widely used tool for evaluation of amusia which only tests contour, interval, scale, rhythm, metre, and musical memory.

In order to be more comparable to speech, all music melodies were monophonic. However, duration was not identical across speech and music stimuli as, apart from the first task that did not have a music counterpart, all the rest of the speech tasks included lexical stimuli. As speech, in contrast to music, has a lexico-semantic dimension, the way one processes speech stimuli is affected by this dimension which is often argued to facilitate processing (e.g., Patel et al., 2008, Liu et al., 2010). For this reason, all music stimuli were longer than the linguistic stimuli in their counterpart tasks. The difference between linguistic and musical stimuli varied across tasks, depending on the melodies and utterances that were chosen for each perceptual process.

In the 'expressive speech prosody' and the 'expressive music prosody' tasks the difference was not as large as in some other tasks. More specifically, in the 'expressive speech prosody' task, the linguistic stimuli had an average duration of 7.93 seconds and the 'expressive music prosody task' included stimuli of an average duration of 8.84 seconds. The difference was much larger across the tasks looking at emotion. The 'emotional speech prosody' task had stimuli with an average duration of 2 seconds, whereas the 'emotional music prosody task' included stimuli of an average duration of 5.84 seconds. The 'speech rate' task and the 'music tempo' task had stimuli that displayed large length differences. The speech stimuli lasted 2.31 seconds in average, while the music stimuli had an average duration of 6.46 seconds.

With respect to the manipulations applied to the stimuli across domains, they were applied in order for the acoustic result to resemble –as much as possible– natural instances of speech and music. This principle was particularly pronounced for the expressive music prosody task and the expressive speech prosody task that were used to test related to the most important research question of the study. The technical manipulation of music stimuli that led to deadpan mechanical melodies resembles old mobile phone ringtones or repetitive riffs in popular music. By contrast, it is less common to hear people produce an analogous type of speech. There are strong reasons why manipulations using synthetic speech might be inappropriate for investigating prosodic features of everyday speech (Vaissière, 2008). Here, the decision was made to ask the speaker who participated in the preparation of the stimuli to create all manipulations orally. This approach was preferred based on this rationale. In general, this methodological hurdle, seeming to be inherent to this comparative endeavour, might warrant further theoretical reflections in designing stimuli and tasks for future research.

The order of presentation of some of the tasks was determined both on the results of some piloting data as well as the nature of the experimental process. Firstly, as it was shown in 3.4, using the first part of the Battery with younger controls showed that the listener can be primed by the lexicalised stimuli of the expressive speech prosody task if these are presented before the speech prosody detection task. For this reason, the speech prosody detection task was the first one to be presented to all participants. Secondly, the two tasks on emotion were shorter than all the other tasks as every question pertained to one stimulus and not two. For this reason, it was thought that the emotion task would have smaller working memory demands and putting them before the optional break and after the first three tasks would make the process easier for participants.

With respect to the development of the Battery designed for this study, it should be noted that an additional task, namely 'music pitch', had initially also been designed and tested on control participants. Participants in both the music experienced and music naïve control groups performed only slightly above chance. Hence, the poor performance of the controls on this task resulted in excluding it from the tasks that were given to IB and BZ and also excluding it from further analysis in general. This was decided, as assessing the musical cognitive profile of a brain-damaged patient and a congenital amusic individual through the use of a task that neurotypical participants were unsuccessful at would not be meaningful However, as these results have methodological implications for the parallel study of language and music, they will be later analysed in the Discussion.

3.2.1 The language of the tasks

The language of the stimuli designed for the speech tasks of the Battery was Greek, as the study was conducted in Greece with monolingual Greek-speaking participants. However, designing the same tasks, for example, in English for English-speaking participants would not have significant differences due to the nature of the prosodic manipulations involved. It is argued below that, with the exception of one acoustic feature which may –to some extent–represent a language-specific variable, the speech tasks of the study would engage the same processes and entail the same degree of difficulty for monolingual speakers of other languages, if they were designed in their mother tongue.

Some speech stimuli which were mainly used in the speech prosody detection task and the expressive speech prosody task included an acoustic feature which does not seem to be present in English. This acoustic feature, which, to the author's knowledge, is not described in the Greek phonology literature, relates to duration lengthening and pitch lowering of one syllable in the last word of an utterance when the speaker tells a story or reaches an important conclusion in narrative discourse. This feature, however, is not likely to have had a significant perceptual effect on the listeners' general performance as it was used only in some of the stimuli in the aforementioned tasks.

As it will be more extensively discussed below, the speech prosody detection task, in general, required the listener's discrimination of pairs of utterances, comparing ones where several features such as pitch , loudness , pausing, and length were varied from a normative version versus utterances where all of these features had been removed. As the above prosodic features are employed in other languages, this task design does not make the relevant processing ability unique to Greek. The same

argument applies to the expressive speech prosody task, where the lexical version of the same stimuli was used.

In the emotional speech prosody task, the use of pitch, length, loudness, and tempo for the creation of happy versus sad stimuli did not reflect any particular properties of Greek. Hence, this task is also argued to be non-specific to Greek.

The speech pitch task tested the perception of a pitch manipulation over a long chunk of utterances. It is believed that the same type of manipulation would require an equal perceptual process in speakers of other languages.

Finally, the speech rate task, as it is later discussed, involved same-different discrimination judgements based on the rate of utterances. This manipulation and its perceptual effect do not seem to depend on language-specific psychological processes.

It is therefore, suggested that the contribution of the study on the linguistic level can be appreciated regardless of a specific language being investigated. As the nature of the manipulations did not capitalise on specific properties of Greek, it can be assumed that similar manipulations in other languages would involve similar processing requirements for non-Greek speaking participants.

3.3 Experimental tasks: the 'LMP'

As the Battery of tasks designed for this study aimed at a parallel exploration of language and music, it will be referred to as LMP (Language and Music Processing Evaluation Battery). New stimuli were recorded and manipulated for the creation of the Battery.

The names of the tasks are as follows: 'speech prosody detection', 'expressive music prosody', 'expressive speech prosody', 'emotional speech prosody', 'emotional music prosody', 'speech pitch', 'speech rate', 'music tempo'.

Table 1 shows the number of stimuli for each task as well as its response mode and the probability of guessing correctly.

Task	Number of stimuli	Response mode	Guessing probability
Speech prosody detection	64	4 answers, forced choice	25%
Expressive music prosody	64	Same-different judgement	50%
Expressive speech prosody	64	Same-different judgement	50%
Emotional speech prosody	32	Identification (2 choices)	50%
Emotional music prosody	32	Identification (2 choices)	50%
Speech pitch	64	Same-different judgement	50%
Speech rate	64	3 answers, forced choice	33.3%
Music tempo	64	3 answers, forced choice	33.3%

Table 1: LMP Battery variables for number of stimuli, response mode, and guessing probability for the 8 tasks.

The following sections give more details on each task separately, explaining the purpose of each task, its format, and the process followed for the creation and manipulation of the stimuli as well as the guidelines provided to the participants during the testing sessions. A detailed description of the instructions that were given to participants can be found in the Appendix.

3.3.1 Speech prosody detection

a) Purpose. This task aimed to investigate participants' ability to discriminate between delexicalised speech stimuli with prosodic inflection and delexicalised speech stimuli in which variation (at the level of pitch, loudness, pausing, and duration) was removed. It was designed in an attempt to give an insight into the participants' ability to detect speech in the absence of lexico-semantic information. That is, it aimed to determine whether variation of the above prosodic cues would be viewed as an indicator of human speech, despite the inability of the listener to identify words.

b) Creation of stimuli. The task consisted of 64 stimuli in total (some of which were repeated in different combinations) presented in pairs. The stimuli had an average duration of 7.93 seconds and, as mentioned earlier, were spoken by a male native speaker of Greek in two conditions: i) exaggerated prosodic inflection ii) minimum possible presence of variation in terms of pitch, loudness, and duration. Half of the pairs included same and half of them different stimuli. For the creation of the prosodically marked speech stimuli, the speaker was instructed to read the sentences with exaggerated pausing between tone-units employing pitch, loudness, and length variation. This intended to add to the aesthetic outcome of the utterances that were designed and make them more interesting acoustically. For the creation of the prosodically unmarked speech stimuli, he was instructed to limit variation in pitch, loudness, and length as much as possible to attempt an imitation of synthetic speech. This principle aimed to make the stimuli aesthetically less interesting for the listener.

c) Manipulations. Praat software (Boersma and Weenink, 1996) was used in order to remove lexical information from all stimuli of this task though low-pass filtering. Low-pass filtering removes frequencies above 500 Hz in a sound spectrum, producing a result resembling speech heard through a wall, that is, semantically

unintelligible but intact in terms of timing and rhythm (Schrerer, 1988). The low-pass filtering technique has been extensively used in the literature to remove lexical information in order to exclude semantic interference. On this task, the stimuli intending to evoke a 'non-speech' answer were the low-pass filtered versions of the stimuli imitating synthetic speech. The stimuli intending to evoke a 'speech' judgement were those generated by the low-pass filtering of the stimuli with exaggerated prosodic variation.

d) Task presentation. This was the first task to be presented to participants. It consisted of 64 stimuli that were presented in pairs.

e) Instructions to the participants. Participants were instructed to attend to each stimulus individually and choose among four possible answers. In order to recognise speech stimuli more easily, participants were given a hint; they were instructed to mark as 'speech stimuli' the stimuli that reminded them of a person speaking behind closed doors.

f) Response mode. The present task included 4 possible answers; the first stimulus is a speech stimulus, the second stimulus is a speech stimulus, both stimuli are speech stimuli, neither of the stimuli is a speech stimulus. Participants had to listen to stimuli carefully and assess them individually. They did not have to compare the stimuli but had, instead, to decide for every of them individually if they reminded them of speech or not. In this task, stimuli were presented in pairs to facilitate identification, although the tasks did not, in fact, require participants to compare the stimuli of each pair. That is, making comparisons throughout the task possibly provided some hints to participants in order for them to discriminate more easily the

two types of stimuli but it was also possible for them to make successful judgements listening closely to each stimulus in isolation.

g) A comment on the task. In the debriefing period, no participants reported having identified the source of the stimuli. This shows that adapting the order of the first tasks after the initial piloting resulted in the improvement of the battery.

3.3.2 Expressive music prosody

a) Purpose. This task was designed to assess ability to distinguish stimuli differing in terms of musical expression in performance. In combination with the expressive speech prosody task (presented in 3.8.3), it also aimed to examine if successful performance on the one task would entail successful performance on the other.

b) Creation of stimuli. The melodies employed for this task were derived from a pool of classical, romantic, late romantic, 20th century, and novel melodies composed by the researcher. All melodies were monophonic adaptations of the original melodies. The pool included different musical styles in order to ensure that the particular character of a single style would not produce any confounds which could affect the judgement of the participants. The stimuli were played by the researcher on a digital piano with weighted keys and melodies were recorded using Digital Performer 8. The melodies of this task had an average duration of 8.48 seconds.

Melodies in each pair of stimuli had identical tune and meter, but half of the pairs displayed differences in terms of the features discussed below. More specifically, as far as deviation from temporal regularity is concerned, the 'rubato' technique was used. This technique allows the performer to deviate from the exact length of notes both by lengthening and shortening their original duration for expressive purposes.

Temporal regularity in the 'expressive' condition was also violated by 'ritardando'; progressive slowing of tempo at the end of a phrase and 'accelerando'; gradual acceleration of time.

Hence, there were two groups of stimuli, hereafter 'expressive' and 'non-expressive'. In the 'expressive' group, melodies displayed variation in dynamics (amplitude differences across the melody), deviation from temporal regularity, and legato articulation (connected and smooth transition between successive notes). The 'non-expressive' group consisted of melodies with temporal regularity, absence of amplitude variation, and marcato articulation (notes were heavily accented). In addition, stimuli belonging to these two categories differed in terms of grouping. 'Expressive' stimuli had small grouping patterns making up the whole melody, whereas 'non-expressive' stimuli did not include these patterns. Half of the pairs differed in terms of the above features and half of them were identical.

c) Manipulations. Initially, the 'expressive' stimuli were recorded and then the 'non-expressive' melodies were generated using commands of the Digital Performer 8 which changed articulation and removed amplitude variation, rhythmic freedom, and pausing.

As mentioned above, articulation was a distinctive feature between 'expressive' and 'non-expressive' melodies. In the 'non-expressive' condition, all notes in melodies had exactly the same 'marcato' articulation. The 'marcato' articulation was technically generated and it intended to add a computerised effect to the melodies. The legato articulation of the 'expressive' stimuli did not have the homogeneous effect of the 'marcato' articulation in the 'non-expressive' melodies. Although

'legato' articulation was prevailing in the 'expressive' melodies, not every note was articulated in exactly the same way as in the case of 'non-expressive' melodies.

Differences in dynamics also contributed to the distinction between 'expressive' and 'non-expressive' melodies. Melodies in the 'non-expressive' condition did not display variation in dynamics. Digital Performer 8 was used in order to apply exactly the same loudness on all notes of a melody, in contrast to the 'expressive melodies' that displayed variation in dynamics according to phrase contours, harmonic relationships, and repetition of motives.

Rhythmic freedom was removed from the non-expressive stimuli on Digital Performer 8, assigning the exact value of every note according to the music score. The same process was followed for rests of non-expressive stimuli that, in contrast to the expressive condition, were kept exactly equal to the values indicated by the score.

d) Task presentation. This was the second task to be presented to participants. Its 64 stimuli were also presented in pairs.

e) Instructions to the participants. Participants were instructed to attend to the stimuli deciding whether melodies were played by a musician or a machine.

f) Response mode. The response mode here required same-different discrimination judgements. If both melodies in a pair were played by a musician, participants would have to indicate 'same'. Likewise, 'same' was the correct answer if both stimuli were played by a machine. 'Different' was the second possible answer if participants judged one melody of the pair to be played by a musician and the other to be played by a machine.

g) Comments on the task. The melodies recorded for this task had to be monophonic, in order to be more comparable to the speech stimuli of the expressive speech prosody task. Moreover, the degree of difficulty was not necessarily the same across melodies, mainly due to the variety of music styles included in the task.

3.3.3 Expressive speech prosody

a) Purpose. This task aimed to explore the ability of participants to discriminate between speech stimuli with total absence of prosodic inflection and speech stimuli with pitch, loudness, pausing, and duration variation. Moreover, as already mentioned, it was designed in parallel with the expressive music prosody task in order to detect the relationship between the two tasks.

b) Creation of stimuli. The stimuli used in this task were the same used in the speech prosody detection task (see description in 3.6.1). The only difference in this task was that the stimuli were used in their lexicalised version, i.e., before applying the low-pass filtering manipulation.

c) Task presentation. This was the third task that participants were presented with. Its 64 stimuli were presented to the participants in pairs.

d) Instructions to participants. The guidelines given to participants for this task differed from the 'speech detection' task, despite the fact that the source of the stimuli was the same. Participants were informed that they would be exposed to recordings of an actor uttering phrases in pairs and they should decide if the actor was equally expressive in generating the sentences of each pair. These guidelines were not given in the speech prosody detection task that included the same stimuli in

a delexicalised form and in that task the listener did not know that all stimuli were originally derived from speech.

e) Response mode. For this task, participants had to make same-different discrimination judgments as in the expressive music prosody task.

3.3.4 Emotional speech prosody

a) Purpose. The emotional speech prosody task looked at participants' perception of emotional prosody, using 'happy' and 'sad' speech stimuli with neutral semantic content. Only two emotions were included in this task, as investigation of perception of emotion was not among the principal objectives of this study. The purpose of including this set of stimuli here was to detect patterns of associations or dissociations between perception of emotion and expressive prosody, and as such does not represent a thorough-going investigation of emotional prosody perception.

b) Creation of stimuli. This task only included 32 stimuli with an average duration of 2 seconds. They were spoken by a Greek male native speaker in the two emotional tones. Half of the stimuli corresponded to happy and half of them to sad emotional tones. Evoking these emotions exclusively depended on the prosodic choices of the speaker, as the semantic content of all sentences was neutral. The speaker produced slow versus fast utterances for the sad condition and his pitch also tended to be lower in the sad condition. In some cases, sad stimuli also were softer and had a slightly lower speaking rate.

c) Task presentation. This 32-stimuli task was the fourth task to be presented to participants. On this task, stimuli were not presented in pairs. Rather, every number in the answer sheet corresponded to a single stimulus.

d) Instructions to participants. Participants were informed that the lexical content of the sentences was neutral and that they were not expected to base their judgement on the meaning of the utterances. Rather, they were instructed to attend to the stimuli focusing on whether the speaker's voice sounded happy or sad.

e) Response mode. For this task, participants did not compare the stimuli. The response mode was identification between two possible choices. For every stimulus, participants had to indicate if it sounded 'happy' or 'sad'.

3.3.5 Emotional music prosody

a) Purpose. This task aimed to explore participants' perception of emotional character in music. As the previous task on emotional speech prosody, this task was not an original contribution to emotion perception. It was primarily designed in order to provide a comparison between perception of 'expressiveness' in music and emotion in music. Hence, although its content was not based on a novel idea, it served a novel research objective.

b) Creation of stimuli. The task consisted of 32 novel melodies of an average duration of 5.84 seconds. These were composed and played by the researcher on a digital piano with weighted keys and all melodies were recorded using Digital Performer 8. In this task, stimuli intending to evoke a happy emotion had a faster tempo, higher loudness, and were composed in a major mode, often also including quick staccato motives. The melodies intending to convey a sad emotion had a slower tempo and lower loudness. They were composed in minor mode and their articulation was different (usually legato). The stimuli were designed in this way, as it was hypothesised that discrimination judgements would be mainly based on slow

versus fast tempo, minor versus major scale, staccato versus legato articulation (in some cases), low versus high loudness, and lower versus higher pitch. As in the relevant speech task, half of the stimuli intended to convey a happy emotion and half a sad emotion.

c) Task presentation. This was the fifth task participants were presented with. As in the previous task, every number in the participants' answer sheet corresponded to one of the total 32 stimuli.

d) Instructions to the participants. Participants were instructed to attend to the stimuli and decide whether the melody they were exposed to sounded happy or sad.

e) Response mode. Similar to the previous task for emotional prosody in speech, participants had to identify the intended emotion. They had to indicate if the stimuli were happy or sad without comparing between stimuli.

3.3.6 Speech pitch

a) Purpose. This task was designed as an alternative method of investigating perception of pitch manipulations on long chunks of spoken stimuli. It was designed to explore perception of pitch differences across spoken chunks of 4 seconds in average.

b) Creation of stimuli and manipulations. The task consisted of 64 spoken stimuli with an average duration of 8.17 seconds. The stimuli for this task were derived from the expressive speech prosody task pool. Half of the pairs included identical stimuli and the rest pairs included some stimuli with a pitch manipulation. This manipulation was applied to the second half of the utterance. More specifically, the

second half of these utterances was transposed two semitones either higher or lower compared to the original stimulus, using the effect 'change pitch' in Audacity software. Every target pair participants were presented with included stimuli with identical semantic content and participants were informed that this would be the case across all pairs.

c) Task presentation. This was the sixth task that was distributed to participants and its 64 stimuli were presented in pairs. In these pairs, the manipulated stimulus was presented first in half of the stimuli and second in the rest.

d) Instructions to participants. Participants were instructed to attend carefully to the second half of each stimulus in order to compare successfully the stimuli. They were also instructed to consider two stimuli as identical only when the whole two utterances of each pair were exactly the same. Finally, they were informed that no semantic differences would ever occur between the utterances of a pair and that they could completely ignore the lexical content of the utterances.

e) Response mode. The response mode of this task required same-different discrimination judgements.

3.3.7 Speech rate

a) Purpose. This task aimed to evaluate participants' rate difference perception in speech, that is, to determine whether participants would be able to differentiate two short utterances on the basis of a speed difference across the total length of stimuli. The task was designed as an additional part of the battery in order to detect possible associations and dissociations in performance compared to other tasks but it is not directly related to the main scope of the present study.

b) Creation of stimuli and manipulations. The task included 64 stimuli that were shortened versions of previous stimuli used in the battery and their new average duration was 2.31 seconds. In 16 target pairs, the stimuli of each pair had the same tempo and, in the rest 16, the stimuli had a 60% tempo difference. This manipulation was applied using the command 'change tempo without changing pitch' on Audacity software. There was no stimulus to which a slow manipulation was applied. All features except for tempo were identical in stimuli of each pair.

c) Task presentation. This was the penultimate task to be given to participants. All stimuli were presented in pairs. In half of the pairs including different stimuli, the faster stimulus was presented first. In the rest of these pairs, the faster stimulus was the second stimulus of the pair.

d) Instructions to the participants. Participants were informed about the kind of comparison they were required to perform for this task, that is, that some utterances would be faster than others in some pairs. As stimuli were short, participants were encouraged to attend carefully to the utterances of each pair in order to be successful.

e) Response mode. The response mode was forced choice among three answers; 'first utterance was faster', 'second utterance was faster', 'both utterances were equally fast'. Hence, for this task, participants did not only need to detect a difference between the stimuli of each pair but also had to specify which of the two stimuli was faster, in case there was a tempo difference.

3.3.8 Music tempo

a) Purpose. This task aimed to assess participants' ability to compare melodies on the basis of a tempo difference.

b) Creation of stimuli and manipulations. A total of 64 melodies with an average duration of 6.46 seconds were included in this task. Some of the melodies were also used in previous tasks of the battery and some others were recorded for this task. Exactly the same manipulation described in 3.6.7 was used for the stimuli of this task. As in the previous task, there were no stimuli that were manipulated to be slower than the original ones and tempo was the only differentiating feature across stimuli. There were 16 pairs in which melodies were of the same tempo and another 16 in which there was a 60% tempo difference.

c) Task presentation. This was the last task of the battery participants were presented with. Half of the different pairs had the fast melody presented first and the rest had the fast melody presented second.

d) Instructions to the participants. Participants were instructed to attend closely to the stimuli and judge if the melodies were equally fast or not.

e) Response mode. As in the previous task, participants had to choose among three answers; 'first melody was faster', 'second melody was faster', 'the melodies were equally fast'.

3.4 Originality, purpose, and more details on the generation and presentation of the LMP tasks

The degree and 'nature' of originality of the LMP Battery varied among tasks. Two of the tasks (expressive speech prosody task and expressive music prosody task) were based on an original testing hypothesis and original stimuli were also created. The speech prosody detection task used a technique which has been widely used in research in order to remove lexical information from linguistic stimuli. Using this

technique in order to delexicalise spoken utterances was not novel. However, using the same technique with spoken utterances of minimum pitch, loudness, and length variation and absence of pausing in order to create stimuli that resulted in sounds that do not resemble speech is novel. The tasks on emotion were not novel in the sense that emotional prosody and emotion in music have been extensively studied before. The stimuli themselves were, however, designed and recorded for this study. The speech pitch task was based on a novel testing hypothesis, as it did not test perception of isolated pitch events but rather examined perception of a pitch transposition of a large chunk of spoken utterances. The speech rate and music tempo tasks were not based on an original research concept but, to the author's knowledge, no tasks in previous research have examined perception of rate in a parallel language-music fashion.

The purpose of the tasks was to investigate how different perception abilities across domains associated or dissociated in the presence of extensive right brain damage or congenital amusia. In addition, as mentioned in the previous section, these tasks also aimed to provide an alternative way of exploring new acoustic features in the parallel approach of language and music.

For the generation of the stimuli employed in this study, Digital Performer 8 was used. Stimuli were recorded in WAV format, 44.1 kHz, 16 bit in a quiet room. All tasks apart from the emotional speech prosody task and the emotional music prosody task included 32 pairs of stimuli. The emotional speech prosody task and the emotional music prosody task did not include pairs of stimuli but rather 32 stimuli, one for each question. A total of 512 stimuli were included in the above tasks. This number of stimuli includes the parent stimuli, their manipulated versions or their repetition depending on the task. Manipulation of sound stimuli for specific purposes

was achieved through the use of Praat software speech analysis (Paul Boersma and David Weenink, 1996), Audacity software (http://audacity.sourceforge.net/) and also Digital performer 8. These manipulations will be detailed below by task.

The presentation and interval between all test stimuli followed the same protocol. In presenting each stimuli, a warning tone preceded each stimulus or pair of stimuli, depending on the task. The stimuli of every pair were separated by one second.

3.5 Participants

3.5.1 Healthy control participants

Greek control participants were recruited through the musical communities and other art communities by word of mouth. Their subject variables for age, education, handedness and socioeconomic status were matched to IB and not BZ (for reasons that will become evident below). The musical experience of IB was also matched for approximately half of the controls.

As part of the preliminary procedure, some screening questions focusing on information about the participants' language learning and musical education background were asked. This aimed to ensure that participants would meet the inclusion criteria for being monolingual and to determine their level of music experience. Exclusion criteria were a history of neurological or psychiatric illness or hearing impairment. Their age, education, and handedness profile was matched with IB's profile. More specifically, their age range was 45-60 years old, they were all right handed, and all had 12 years of education. Their socioeconomic status corresponded to 'skilled manual workers' or 'administrative jobs'.

With respect to the musical experience of the control participants, 14 of them had spent 2 to 4 years either singing in a choir or playing a musical instrument as did IB and another 10 were music naïve. The control participants were tested in Chios and Thessaloniki, Greece in a quiet room in their house.

3.5.2 Brain-damaged participant: IB

The brain-damaged participant of the study is referred to as IB. He is a 51 year old Greek monolingual right handed man with 12 years of education and 3 years of music experience as an amateur drummer in a musical band who worked as a bus driver and president of a football club. He had normal cognitive development and no history of mental illness. He also had normal hearing and vision.

He was admitted to Renaissance Rehabilitation Centre on 21/08/2013, in Thessaloniki, Greece after suffering a stroke on 31/07/2013 resulting in left hemiplegia. He had previously been admitted to another clinic between 31/07/2013 and 21/08/2013. He manifested no aphasic symptoms, visual or hearing impairments. He seemed aware of his condition. He also initially manifested some exaggerated negative emotions, such as fear and impatience.

With respect to IB's lesion site, five CT scans were obtained 1 and 9 months post-stroke. A scan taken on 2/8/13 showed an extended ischemic infarction of the right middle cerebral artery territory with mild pressure phenomena in the right ventricle. An additional scan four days later revealed a similar clinical picture. A scan taken on 16/8/13 showed a reduction of the extent of the initially damaged areas. Only limited bleeding was observed. The pressure phenomena in the right ventricle were also reduced. A normal blood flow of the main arterial stems was also observed. The last

scan was taken 7 months later. The scan depicted an old extended ischemic infarction of the right middle cerebral artery territory with some attraction of the right lateral ventricle. The brain stem and the cerebellum were normally depicted.

On admission, IBs' general neuropsychological functioning was assessed with a Mini Mental State Examination. This was the only tool that was used by the clinicians in order to evaluate his higher cognitive functioning. The patient scored 26 out of 30 on this test. The test included orientation questions, word registration (repeating words after a short pause), attention and calculation, recall, repetition, word naming, three stage commands, shape copying, and performance of a written command. When the test was repeated a month later, IB scored 29 out of 30 questions correct.

IB received a certain amount of facial muscle therapy. As his speech production and perception was intact, the speech therapists focused on exercises to restore his facial muscle strength. As he did not display any linguistic or pragmatic impairment, he did not receive rehabilitation for higher cognitive functions. He received therapy for his sensory-motor impairments, including physiotherapy sessions twice or three times per week and exercises in a swimming pool twice a week. These sessions aimed to restore his sensory-motor deficits subsequent to his left hemiplegia.

For the present investigation, testing of IB was carried out by the author in Renaissance Rehabilitation Centre 11 months post onset of illness under the supervision of Dr Harriet Proios, PhD CCC-SLP, Head of the Speech Therapy Department at the Renaissance Rehab Center and Assistant Professor of Neurocognitive Disorders and Rehabilitation at the Department of Educational and Social Policy, University of Macedonia, Thessaloniki, Greece. He was also very

willing to co-operate with the examiner. Due to becoming easily fatigued, however, testing was completed in a number of short sessions.

Baseline auditory perceptual screening of IB

IB was tested on some baseline tasks involving perceptual judgments of basic aspects of auditory stimuli to ascertain his fitness to be tested on the more complex experimental tasks which form the focus of the present investigation. These tasks involved comparisons of auditory stimuli of short duration. More specifically, the baseline stimuli were single tones and Greek vowels that required same-different discrimination judgements. Different pairs were created using Audacity software (http://audacity.sourceforge.net/). There were three categories of stimuli for both vowels and tones; pitch, duration, and loudness. Pitch stimuli differed in four semitones, duration and loudness stimuli also had a difference easily perceived by the normal listener.

The music pitch baseline task included stimuli in pairs that either differed by 4 semitones or were identical. The analogous speech task included the stimuli of Greek vowels (a, e, i , o, u) that were either identical in pitch or had undergone a manipulation of a 4 semitone difference. The duration baseline stimuli across domains tested IB's ability to discriminate between music notes or vowels that either had identical duration (1 second) or differed by 1 second (1 stimulus lasting for 1 second and the other one lasting for 2). The loudness baseline test included stimuli of default loudness (same as in the previous baseline tests) and others of louder vowels or notes, using the command 'effect: amplify' in Audacity software).

These tasks aimed to assess IB's ability to compare simple speech and music stimuli that were not demanding in terms of attention or cognitive load. These baseline tasks were designed as a means to investigate IB's low-level processing and its relationship with higher level processing of similar acoustic stimuli. IB performed successfully on these tasks and made a very small number of mistakes. He scored 7/10 on the amplitude test in music, 8/10 on the amplitude test in speech, 8/10 both on the music pitch and the speech pitch test, 8/10 on the duration test in music and 9/10 on the duration task in speech. These results indicate that his low-level processing was not impaired. It was clear that he could understand what was required and sustain attention and engage in this type of testing. Moreover, IB demonstrated that, although he had suffered extensive brain damage, he was not impaired on tasks of simple auditory judgements which might be argued to underpin more complex discriminations.

Hence, given the fact that he was successful on these tasks, his performance on the main experimental tasks of this study would not be attributed to low-level pitch, duration, or loudness deficits.

3.5.3 Congenital amusic individual: BZ

BZ is a Greek congenitally amusic individual that was originally studied in Paraskevopoulos et al. (2010). BZ is a right handed Greek mother tongue multilingual woman now aged 68 years old with a MD and PhD in Medical Sciences. BZ did not have any positive history of neurological deficit but reported that she had difficulty in remembering music tunes and dancing to the rhythm. Tests on a series of neurocognitive abilities (working memory, recall, visual memory,

attention, executive function, and word fluency) did not reveal any relevant deficits (Paraskevopoulos et al., 2010).

This earlier study preceding the present investigation had tested BZ on two music cognition batteries; the Montreal Battery of Evaluation of Amusias (MBEA) and the Greek Battery of Evaluation of Amusia (GBEA). The MBEA includes 6 music tests, namely contour, interval, scale, rhythm, metre, and music memory with 30 questions each (Peretz et al., 2003). The MBEA was adapted as the GBEA by Paraskevopoulos and colleagues to Greek/Eastern music in an attempt to determine whether using a Western music battery is appropriate for assessing the musical abilities of participants from different musical traditions. This test was validated with Greek controls.

Paraskevopoulos et al. (2010) found that BZ's scores on the GBEA were much poorer compared to neurotypical participants and hence identified her as being amusic. However, it should be noted that the control group was not matched to BZ in their study. BZ was more highly educated, more than 40 years older compared to the mean age of neurotypical participants, and not necessarily of the same socioeconomic status.

Several years after the publication of their study, the investigator of the present study contacted Paraskevopoulos in order to gain access to BZ. BZ agreed to further testing of her music ability. She was tested on the novel Battery of the present study in Thessaloniki, Greece where she lives.

3.6 Experimental procedure

Stimuli were binaurally presented to all participants through earphones, using Windows Media Player. Participants gave written answers by ticking boxes on the provided answer forms. Loudness was adjusted to a comfortable level for each participant. Before every task, participants were given practice items in order to familiarise themselves with the task format and to ask questions. Participants were given positive feedback during the testing session regardless of their performance.

3.6.1 Testing of control participants

The majority of control participants were tested individually. In a few cases, participants were tested in pairs with the use of USB cable. In these cases, special attention was given so that both participants would receive sufficient guidance in order to complete the task. Average duration for the completion of all tasks was two hours. Participants decided if they wanted to take breaks throughout the process. To guarantee that participants attended to all stimuli, rest breaks were given and, at the same time, participants were required to indicate the number of question they were completing.

3.6.2 Testing of IB

Due to the fact that IB was easily fatigued, he completed the tasks in various sessions, depending on his therapy schedule at the Rehabilitation Centre. The majority of sessions took place in the Rehabilitation Centre and only a couple of sessions took place after IB was discharged home from the Rehabilitation Centre. IB decided on the meeting days and most sessions took place in the late afternoon after he had taken some rest. During the sessions taking place at 'Renaissance', the patient

was seated in his wheelchair and was instructed to attend carefully to the stimuli. He was instructed to inform the examiner in case he felt fatigue or wanted to take a break. Moreover, he was aware of the fact that he was free to stop the task and carry on a different day. Rest breaks were also taken regularly during the process and IB was asked to indicate the number of question he was filling in. In most cases, a task was completed in two separate sessions of approximately 20-30 minutes duration, including delivery of detailed instructions. This ensured that IB attended to all stimuli. In addition, every question number was read aloud either by him or the examiner to ensure that he was following the test.

There was no evidence of any practice effect; IB's success or failure on each task was consistent across sessions of the same task. On tasks where he performed poorly, he had a response bias that favoured the judgement of stimuli as the same.

3.6.3 Testing of BZ

BZ was tested in a quiet room in her house, completing all the tasks in a single session with no breaks. Given her relationship with music and possibly her experience participating in the previous research project (Paraskevopoulos et al., 2010), BZ felt comfortable with the speech tasks but appeared to be less comfortable with the music tasks in general. She explicitly expressed this attitude before the beginning of the first music task, referring to her bad experience with music in the past.

3.7 Ethics

This research project gained approval from the SSHP Ethics Committee of Birkbeck, University of London. In addition, the Committee of Renaissance Rehabilitation Centre also granted ethical approval for this research project. They granted the researcher's access to the participant IB who was a patient at the Centre. The testing protocol and the consent forms were approved by their research board, the Scientific Director, and the President of the Centre.

This research was deemed to present minimal risk for the participants and the potential to cause distress was very limited, as they were not exposed to any unpleasant stimulus or procedure. The researcher protected the physical, social, and psychosocial wellbeing of all participants as well as their confidentiality, obtaining their consent and informing them about the study. The consent form which they all signed was accompanied by an additional sheet providing them with information about the study in simple language. Hence, participants were aware of the fact that they had the right to withdraw from the project if they wished so. Moreover, contacting information of both the researcher and the Professor supervising the project was provided. The information obtained by the participants remained confidential by assigning identification codes to them. Any identifying information was kept separate from the data.

To make sure that participants would not be fatigued, they were informed that they could take regular breaks if they wished so. When the brain-damaged individual completed the tasks, testing sessions were kept short. All participants were given positive feedback regardless of their performance to promote a non-stressful experience for everyone.

4. Data Analysis and Results

This chapter presents the results that were obtained from control participants as well as IB, and BZ. The following section initially presents an overall comparison of the performance of two control groups for the music experience variable. In a later section, the results from the performance on controls are presented on each task in the LMP Battery in more detail. Performance of IB is also presented relative to the performance of controls. A similar process is followed for the presentation of results on BZ's performance. The reader, however, has to bear in mind that controls are not matched with BZ and the performance of healthy individuals are presented in parallel as a simple reference point for the appreciation of BZ's performance.

4.1 Control participants: the variable of musical experience

As mentioned above, the control group consisted of two smaller groups differing on a single variable. This variable had to do with whether participants had musical experience or not. A group with musical experience was thought to be crucial, as the brain-damaged individual to whom control participants were matched had had 3 years of music experience. However, participants with no musical experience were also recruited in order to detect possible differences between groups differing on that very relevant variable.

Statistical analysis was conducted in order to determine if the variable for musical experience was statistically significant in the performance of control participants. The total number of controls was 24 and 14 of them had musical experience matched with IB; the other 10 participants were musically uneducated. Independent t-tests were carried out on the performances of the two control groups: Musically naïve

group n= 10 and musically experienced group n= 14 on 8 tasks. No statistically significant difference was found between the two groups on 7 of the tasks: 'Speech prosody detection' t(22)= 1.3, p= .18, 'Expressive music prosody' t(22)= -.17, p= .87, 'Expressive speech prosody' t(22)= -.09, p= .92, 'Emotional speech prosody' t(22)= -.87, p= .39, 'Emotional music prosody' t(22)=-.05, p=.95, 'Music tempo' t(22)=2, p=.06, 'Speech rate' t(22)=-1, p= .32. Only one of the 8 comparisons approached significance: 'Speech pitch' t(22)= -2.4, p= .02. However, examining the scores of both groups on this task, the non-musicians' performance mean score was 30.6 with a range of min-max: 28-32, while musicians' performance mean score was 26.85 with a range of min-max: 18-32. Given that some participants in both groups achieved ceiling performances, while a small number of musicians performed relatively less well, it could not be argued that music experience was a positive variable for this task. In general, given that the two groups could not be demonstrated to differ in performance, their results were collapsed. In the results reported below the control participants will presented as one group N= 24.

4.2 Control Participants as a single group

This section presents the results that were obtained from control participants. Figure 1 gives a picture of the spread of individual controls' scores on all tasks, including the task that was later excluded from further investigation (as mentioned in section 3.5). In Figure 1, the dots of the diagram represent individual scores of all control participants and this display allows for an overall appreciation of controls' performance on the tasks of the Battery.
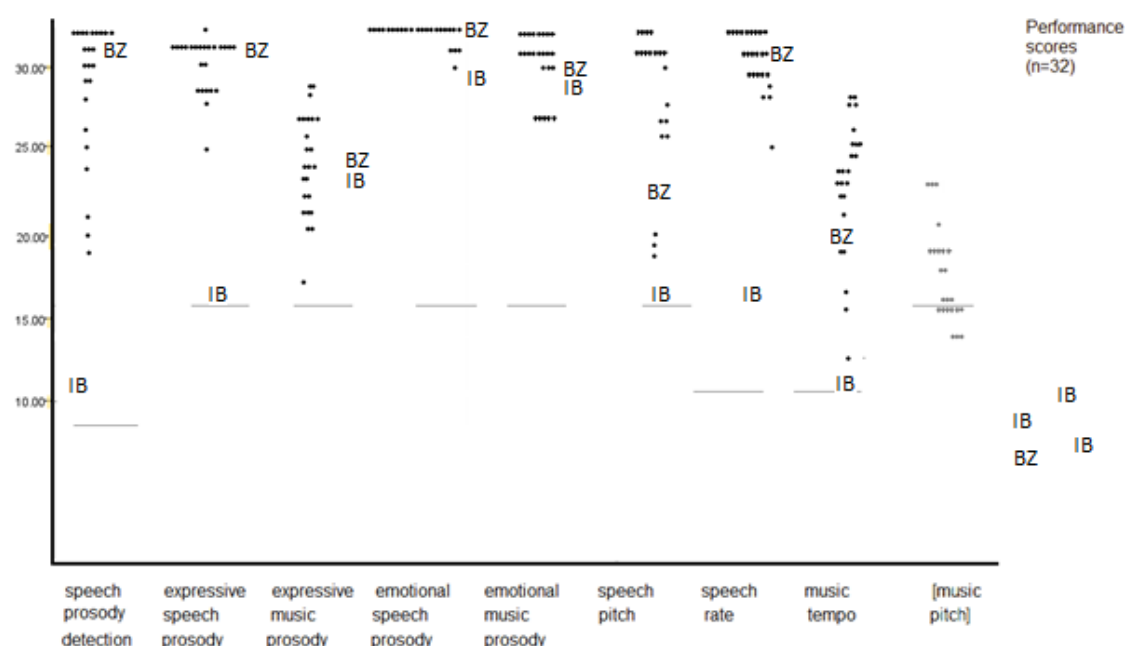
Figure 1: Individual performance scores for 24 control participants, IB, and BZ on the 8 tasks of the LMP Battery plus the music pitch task that was later excluded. The x-axis shows the task in question and the y-axis is measuring the score of participants. Horizontal lines indicate performance at chance.

The results of controls for the 8 tasks of the Battery are presented in Table 2, where

the mean, the standard deviation, and the range of scores are shown.

Tasks	Speech prosody detection	Expressive speech prosody	Expressive music prosody	Emotional speech prosody	Emotional music prosody	Speech pitch	Speech rate	Music tempo
Mean	28.83/32	30.16/32	24.16/32	31.79/32	30.37/32	28.41/32	30.58/32	23.29/32
SD	4.12	1.49	3.15	0.50	1.88	4.09	1.71	4.15
Min-max	19-32	25-32	17-29	30-32	27-32	18-32	25-32	13-29

Table 2: Performance of controls on the 8 tasks of the Battery.

Results from neurotypical controls are presented below. Whisker plots follow the description of single tasks or pairs of analogous speech-music tasks in order to give a schematic representation of the range of performance and of possible outliers.

a) Speech prosody detection. Control groups mean score on this task was 28.83/32 (min-max: 19-32, SD= 4.12). These results suggest that controls were able to discriminate between sequences with pitch, length, and loudness variation and sequences with absence of these features in a delexicalised context. Performance of controls on this task is represented in Figure 2.
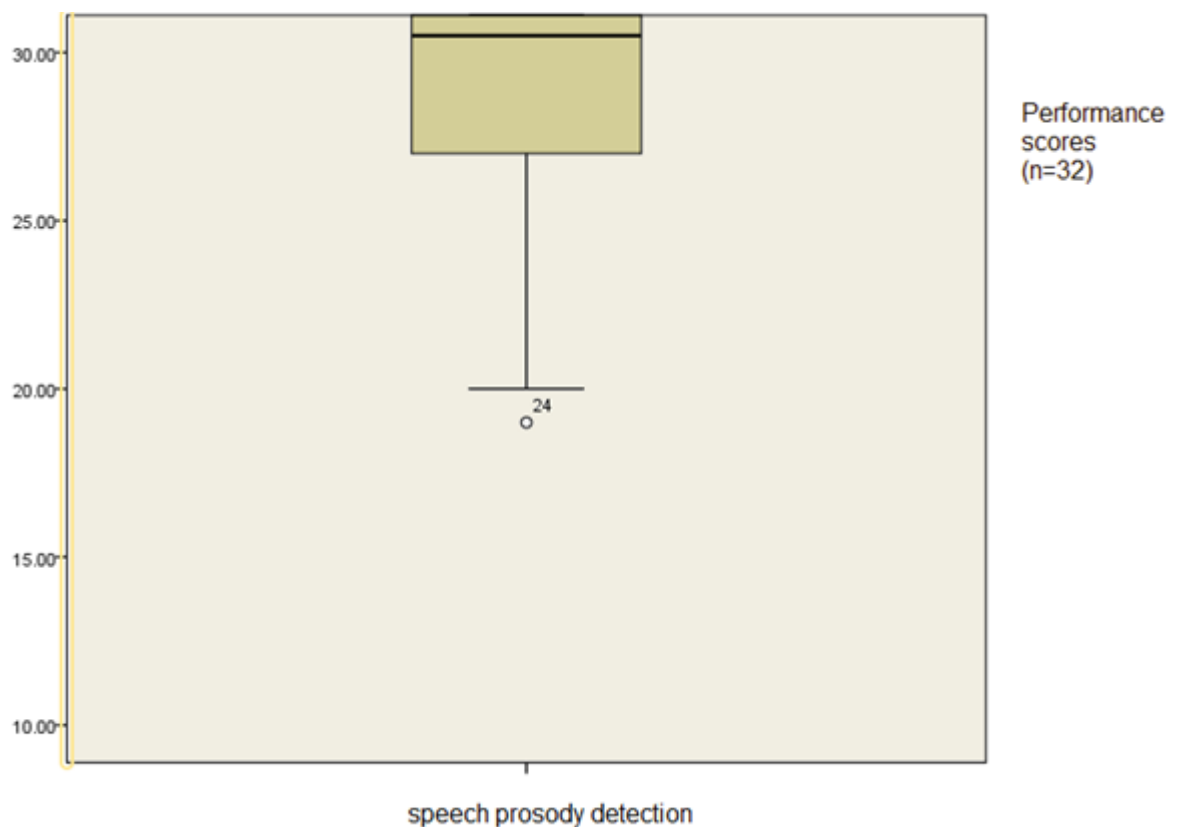

Figure 2: Whisker plot displaying the range of performance of controls on the speech prosody detection task. There is one outlier represented by an open circle indicating that participant number 24 fell outside the range of the other controls. The chance level for this task is a score of 8/32. The horizontal line in the box shows the median. The circle shows an outlier, participant number 24.

b) 'Expressive music prosody. On this task, controls' mean score was 24.16/32 (min-max: 17-29, SD= 3.15). Although their performance was not at ceiling, controls were able to discriminate between expressive and non-expressive melodies.

c) Expressive speech prosody. Control participants scored 30.16/32 (min-max: 25-32, SD= 1.49). This finding suggests that healthy individuals do not have difficulty in discriminating between spoken utterances with differences in pitch, loudness, and length. Although not comparable to the speech prosody detection task due to the different response mode, it can be suggested that the presence of lexical information facilitates discrimination of stimuli that display the above differences.   Performance on this task and its analogous music task is schematically presented in Figure 3.
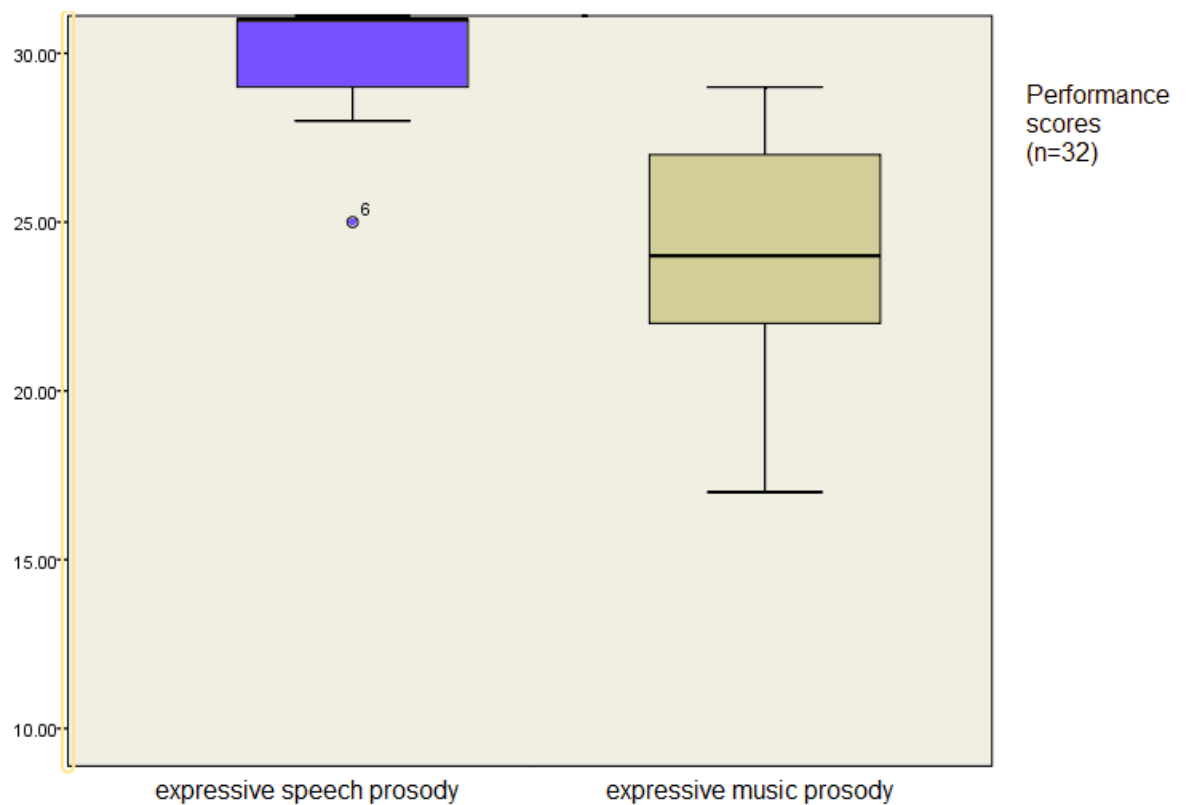
Figure 3: Whisker plot displaying the performance of controls on expressive speech prosody and expressive music prosody. On the expressive speech prosody task, there is one outlier represented by an open circle indicating that participant number 6 fell outside the range of the other control participants. Chance level for both tasks displayed here is a score of 16/32. The horizontal lines in the boxes show the median. In the expressive speech prosody box, the median line almost overlaps with the upper line of the box.

d) Emotional speech prosody. The controls' score on this task was 31.79/32 (min-max: 30-32, SD= 0.50). The high score of participants on this task suggests that the intended emotional tones were successfully conveyed.

e) Emotional music prosody. This task was also very easy for controls. They scored 30.37/32 (min-max: 27-32, SD= 1.88). The high score of all participants suggests that the intended emotion was also successfully conveyed in this task. The slightly better performance of participants on emotion in speech can be possibly attributed to a small number of ambiguous stimuli included in this music task. Controls'

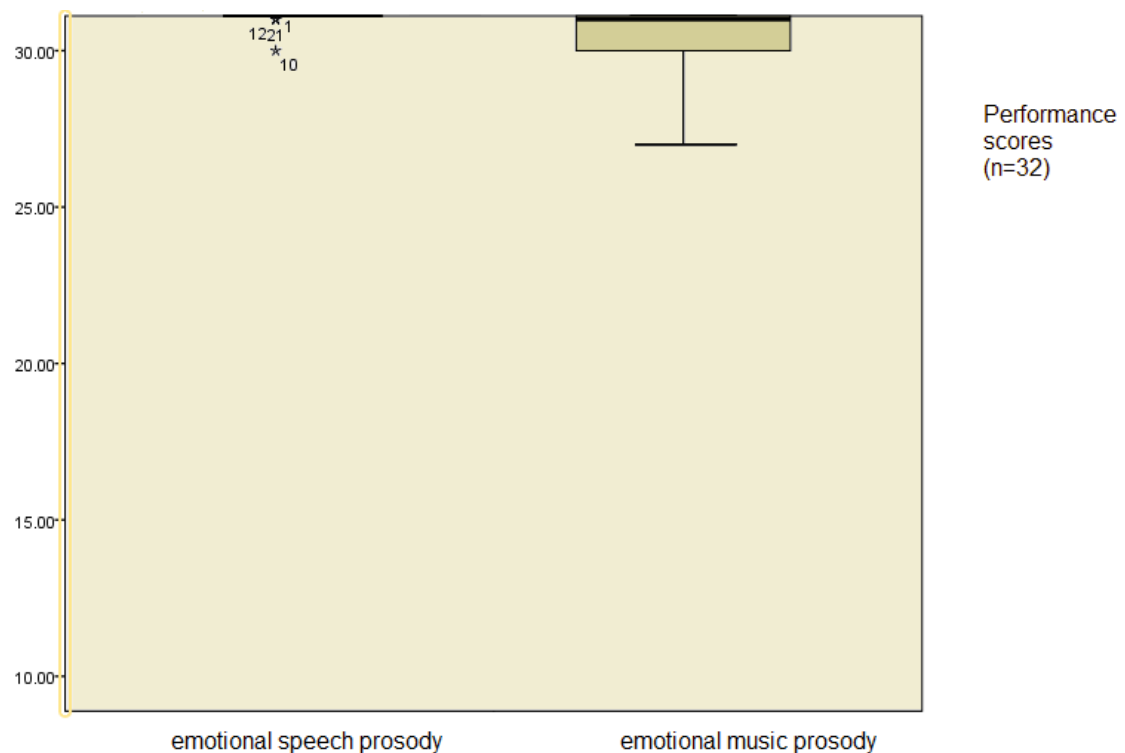performance on the emotional speech prosody and emotional music prosody task is represented in Figure 4.

Figure 4: Whisker plot displaying the range of performance of controls on emotional speech prosody and emotional music prosody tasks. Participants number 1, 10, 12, and 21 were the only controls that did not achieve a score of 32/32. Participant number 10 scored 30/32 and the other three scored 31/32. Dark horizontal lines show the median.

f) Speech pitch. This was a relatively easy task for most controls. Their mean score was 28.41/32 (min-max: 18-32, SD= 4.09). As it is shown in the diagram in this section, only three participants scored low on this task. Even these three participants scored higher than chance as guessing probability was 50%. The other participants scored much higher. It is assumed these controls did not feel comfortable to ask additional questions on the format of this task and the procedure, as 11 participants scored more than 30/32, that is, they made either a single mistake or no mistake at

all. It is possible that differences among participants were due to the format of the task rather than its actual difficulty.

Although this task is based on a pitch manipulation, good performance is likely to be attributed both to perception of pitch but also to perception of voice quality rather than solely pitch perception, as there was a distinctive difference in voice quality. This was an alternative way to test pitch perception, as detecting pitch differences between stimuli differing in one tone, for example, might be deemed harder, especially if the ability to closely attend to a stimulus is impaired. Controls' performance on this task is presented schematically in Figure 5.
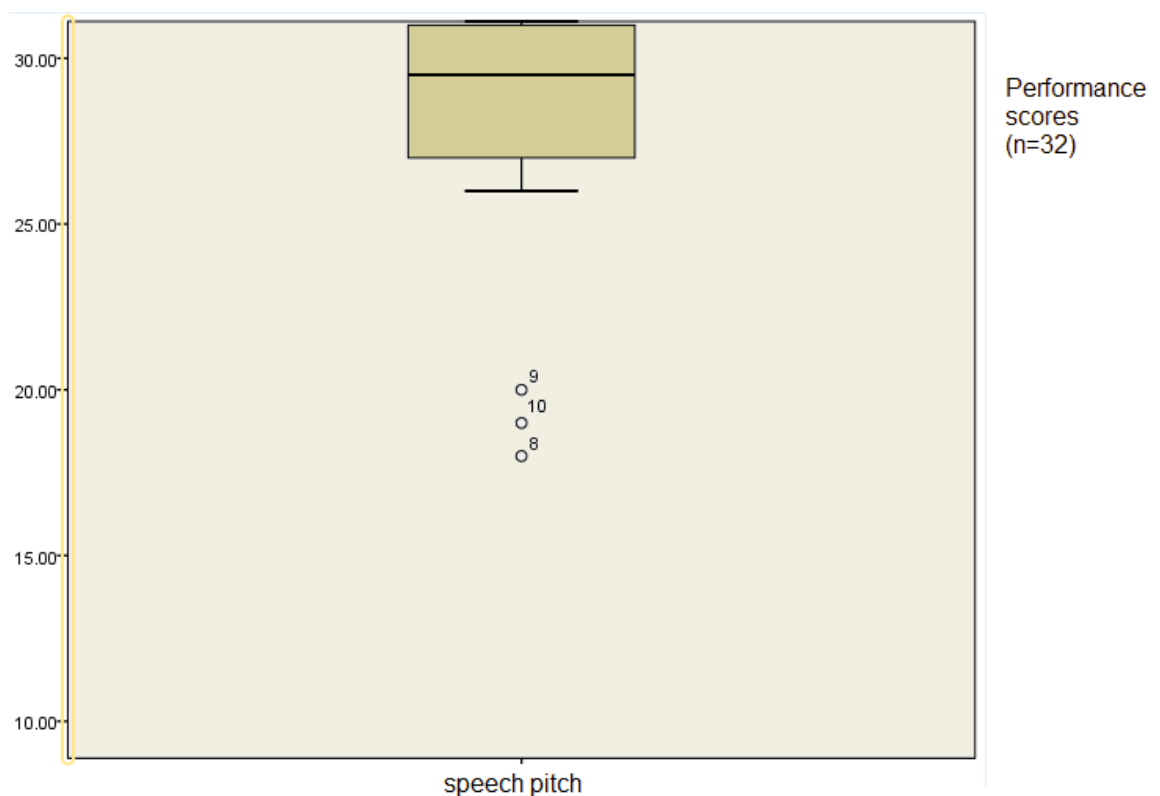

Figure 5: Whisker plot displaying the range of performance of controls on the speech pitch task. Open circles indicate that participants number 8, 9, and 10 fell outside the range of the rest of controls on this task. Chance level score for this task is 16/32. The horizontal line shows the median. Participants number 9, number 10, and number 8 achieved a score much lower compared to the rest of controls.

h) Speech rate. On this task, the control group scored 30.58/32 (min-max: 25-32, SD= 1.71). Their ceiling performance suggests that controls found it very easy to perceive speech rate differences in stimuli with an average duration of 2.31 seconds.

g) Music tempo. Controls here scored 23.29/32 (min-max: 17-29, SD= 4.15). Their score was not very high, however, it was still well above the guessing probability of 33.3% or approximately 10/32. As the response mode for this task was forced choice among 3 possible answers, the probability of guessing correctly was lower compared to the binary same-different discrimination tasks of the LMP Battery. No comparison is attempted here between the 'music tempo task' and the 'speech rate task', as there were large differences in the duration of the stimuli across the domains. However, it is observed that it was easier for controls to compare linguistic rather than music stimuli in terms of overall speed, according to the results of the present tasks. The whisker plots in Figure 6 show performance of controls on the music tempo task and its analogous speech task.
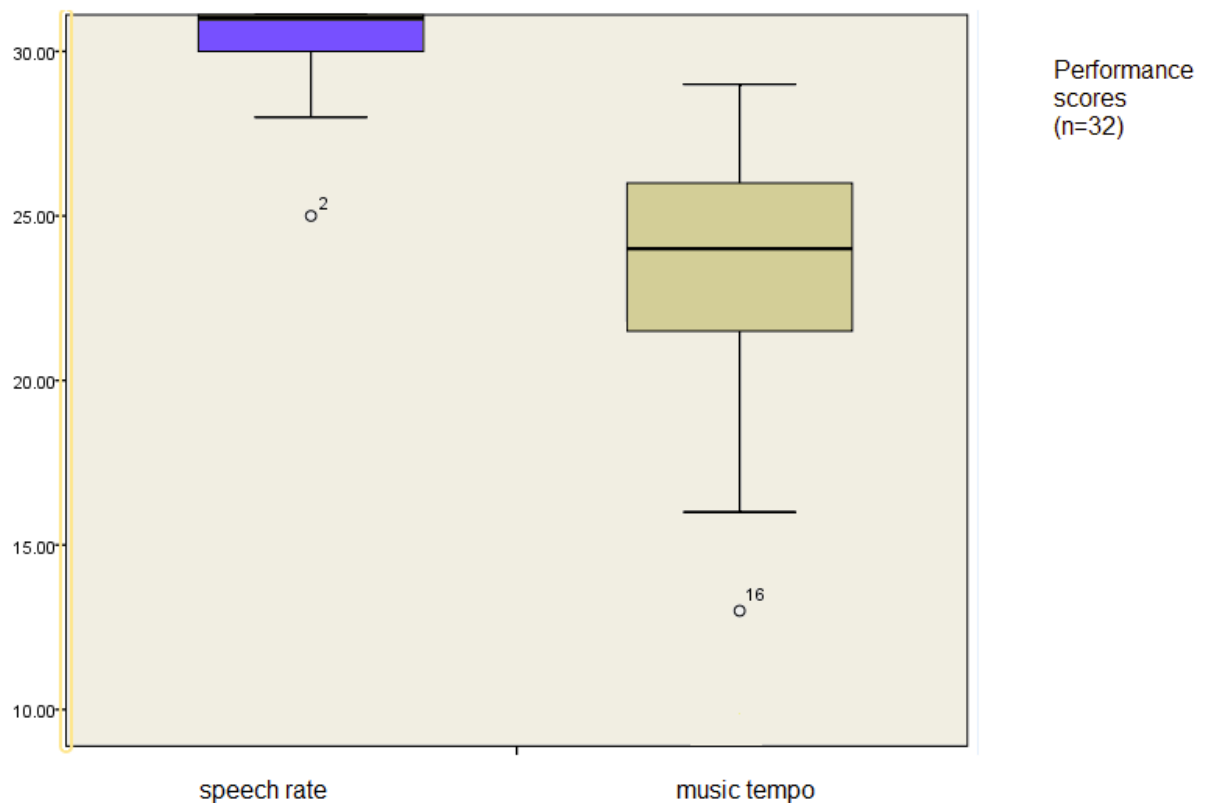
Figure 6: Whisker plot displaying the range of performance of controls on speech rate and music tempo tasks. Participant number 2 and participant number 16 were outliers in the speech rate and music tempo tasks respectively. The chance level for both tasks is a score of 10.7/32. Dark horizontal lines in boxes show the median in each task. Participant number 2 in the speech rate task and participant 16 in the music tempo task scored outside the range of the rest of controls.

The above results suggest that neurotypical individuals do not have difficulty in detecting speech prosody, appreciating expressive performance in speech, and perceiving emotion across domains. They are also good at distinguishing between expressive and non-expressive performance in music. Perceiving transpositions of chunks of an utterance in spoken stimuli also seems easy for healthy individuals. According to the present tasks, perceiving speed differences in spoken stimuli appears an easier task compared to perceiving speed differences in the musical domain.

143

4.3 IB's performance results

In this section, IB's performance on the 8 tasks of the LMP battery is compared to that of neurotypical controls. Whisker plots display his performance in comparison to the controls' performance to represent visually his spared and impaired abilities.

a) Speech prosody detection. On this task, IB's performance was very poor. He scored 11/32. However, while his performance was at chance, his response pattern was not random guessing. fact, IB favoured the option 'both sounds resemble' speech in almost all the questions. These results show that he was unable to successfully discriminate between repetitive monotonous sequences and sequences with rich contour variation and length and loudness variation. It is notable that this was a task on which most controls performed at or near ceiling. Their mean score was 28.83/32 (min-max: 19-32, SD= 4.12). According to Crawford and Howell, (1998) significance test, the difference between IB and controls is statistically significant (t(29) = -4,240; p=0,000, estimated percentage of normal population falling below individuals' score = 0,00%). The performance of IB on this task is presented relative to controls in Figure 7.
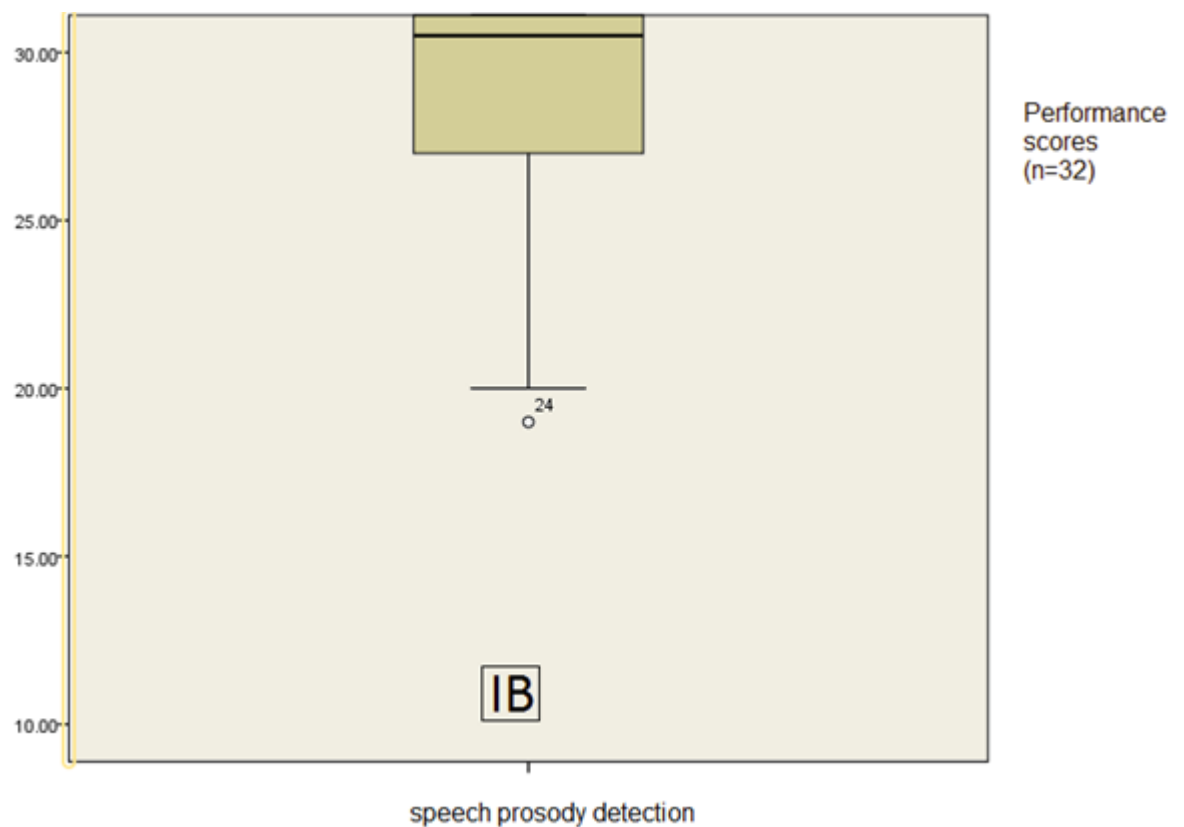
Figure 7: IB's performance relative to controls on the speech prosody detection task. The horizontal line in the box shows the median. The circle shows an outlier, participant number 24.

b) 'Expressive music prosody. IB's performance on the 'expressive music prosody task' was strikingly good. On this same-different discrimination task, IB scored 23/32, outperforming some controls. Recall that controls' mean score was 24.16/32 (min-max: 17-29, SD= 3.15) on this task. According to Crawford and Howell, (1998) significance test, the difference between IB and controls is not statistically significant (t(29) = -0,361; p=0,722, estimated percentage of normal population falling below individuals' score = 36,08%).

c) Expressive speech prosody. On this task, IB's performance was rather poor. His score was 16/32, whereas control participants' mean score was 30.16/32 (min-max:

25-32, SD= 1.49). Controls' high score shows that this task was particularly easy for neurologically normal individuals which emphasises IB's impairment on this task. IB's answers favoured the answer 'both stimuli sound the same'. In fact, at some points, he emphatically uttered 'exactly the same, exactly the same' when performing the task. Results suggest that the presence of linguistic information in this task did not seem to benefit IB's ability to discriminate between the stimuli that were presented in a delexicalised form in the speech prosody detection task. According to Crawford and Howell, (1998) significance test, the difference between IB and controls is statistically significant (t(29) = -9,311; p=0,000, estimated percentage of normal population falling below individuals' score = 0,00%). Figure 8 shows IB's performance relative to controls on the expressive speech prosody task and its analogous music task.
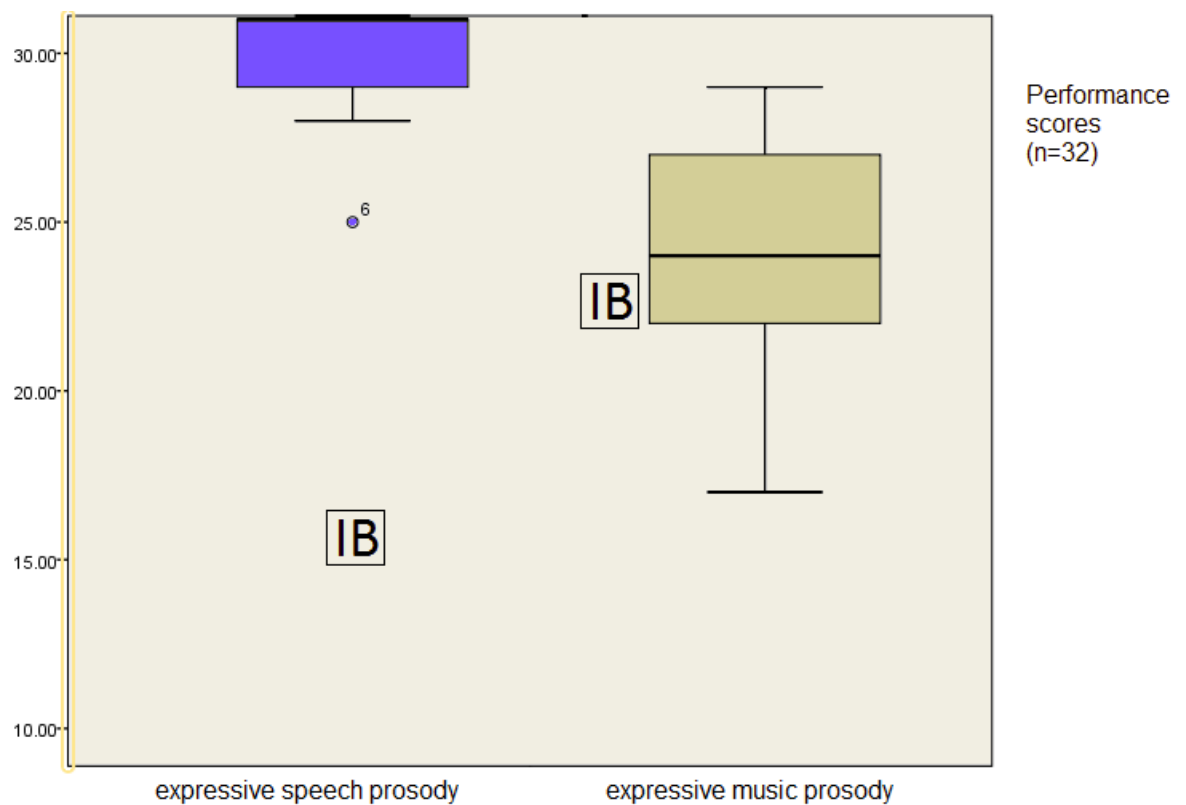
Figure 8: IB's performance relative to controls on the expressive speech prosody and the expressive music prosody tasks. Dark horizontal lines in the boxes show the median. In the expressive speech prosody task, there is one outlier, participant number 6.

d) Emotional speech prosody. IB's performance here was good (29/32) and well above chance. Controls' score was higher 31.79/32 (min-max: 30-32, SD= 0.50). However, the standard deviation of controls was very small with 20 out of 24 participants scoring at ceiling. For this reason, the difference appears to be statistically significant but it can be argued that 3 wrong answers out of 32 can be attributed to IB not attending to these 3 stimuli rather than having impaired emotional speech prosody perception. The results from Crawford and Howell, (1998) test are as follows: $t(29) = -5,467$; $p=0,000$, estimated percentage of normal population falling below individuals' score = 0,00%. If results are considered holistically, perception of happy-sad tones is not proved to be compromised in IB.

This can further suggest that perception of emotional prosody does not seem to rely on pitch perception (which appears impaired in IB according to speech prosody detection task and speech pitch task) or even if it does, other features seem to make up for a possible pitch perception deficit.

e) Emotional music prosody. According to results, this task was also easy for both control participants and IB. IB scored 27/32 and, as mentioned above, controls scored 30.37/32 (min-max: 27-32, SD= 1.88). According to Crawford and Howell, (1998) significance test, the difference between IB and controls is not statistically significant (t(29) = -1,756; p=0,092, estimated percentage of normal population falling below individuals' score = 4,62%). Performance of IB on this task and its analogous speech task is presented in Figure 9.
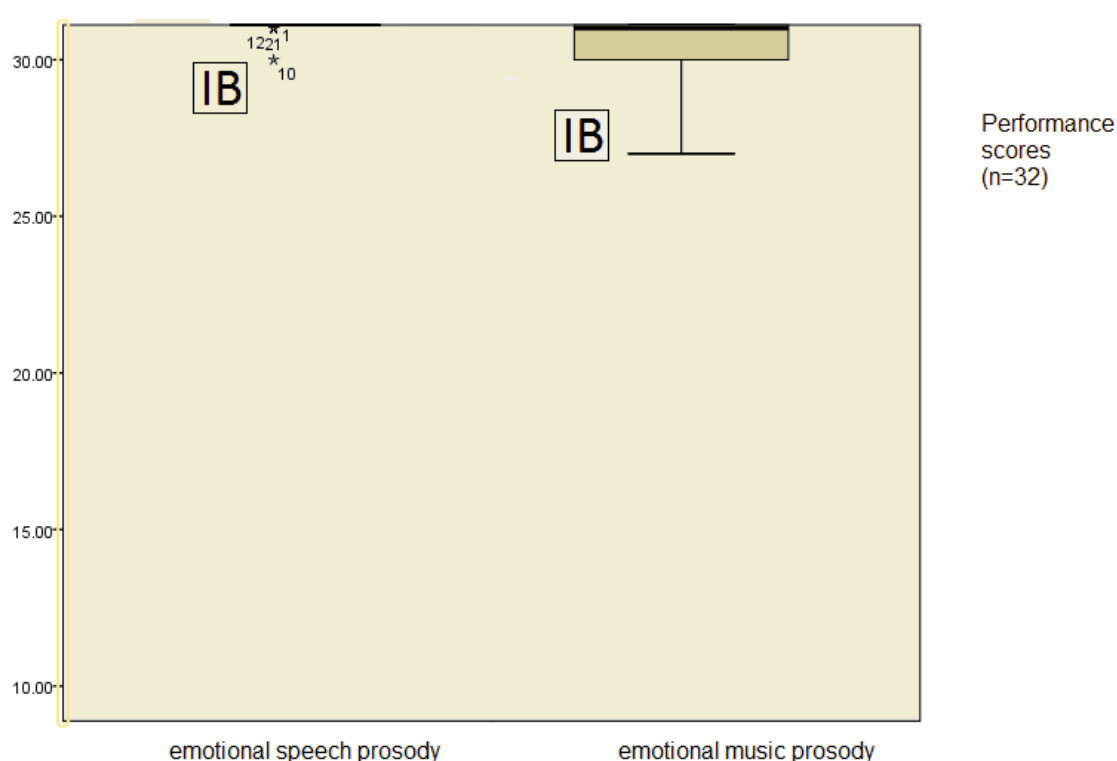

Figure 9: IB's performance relative to controls on the emotional speech prosody and the emotional music prosody tasks. Dark lines show the median for each task. In the emotional speech prosody task, participants marked with a star (1,12,21,10) were the only participants that did not score 32/32. These participants scored 31/32 apart from participant number 10 who scored 30/32.

f) Speech pitch. IB's performance on this same-different discrimination task was at chance (16/32) while controls' mean score was 28.41/32 (min-max: 18-32, SD= 4.09) IB failed to perceive pitch changes that were noticeable by most controls. Some controls did have difficulty with this task, with one scoring only 18/32. As mentioned above, comments by some controls suggest that there might have been a difficulty in understanding the format of this particular task and the response required. However, IB's poor performance on this task cannot be attributed to this factor, as the examiner gave him more detailed information about the format of the task. Moreover, IB was given more time between pairs of stimuli and also the task was completed in two separate sessions. His performance was consistent across sessions despite this procedure. According to Crawford and Howell, (1998) significance test, the difference between IB and controls is statistically significant (t(29) = -2,973; p=0,007, estimated percentage of normal population falling below individuals' score = 0,34%). Figure 10 shows IB's performance relative to controls on the present task.
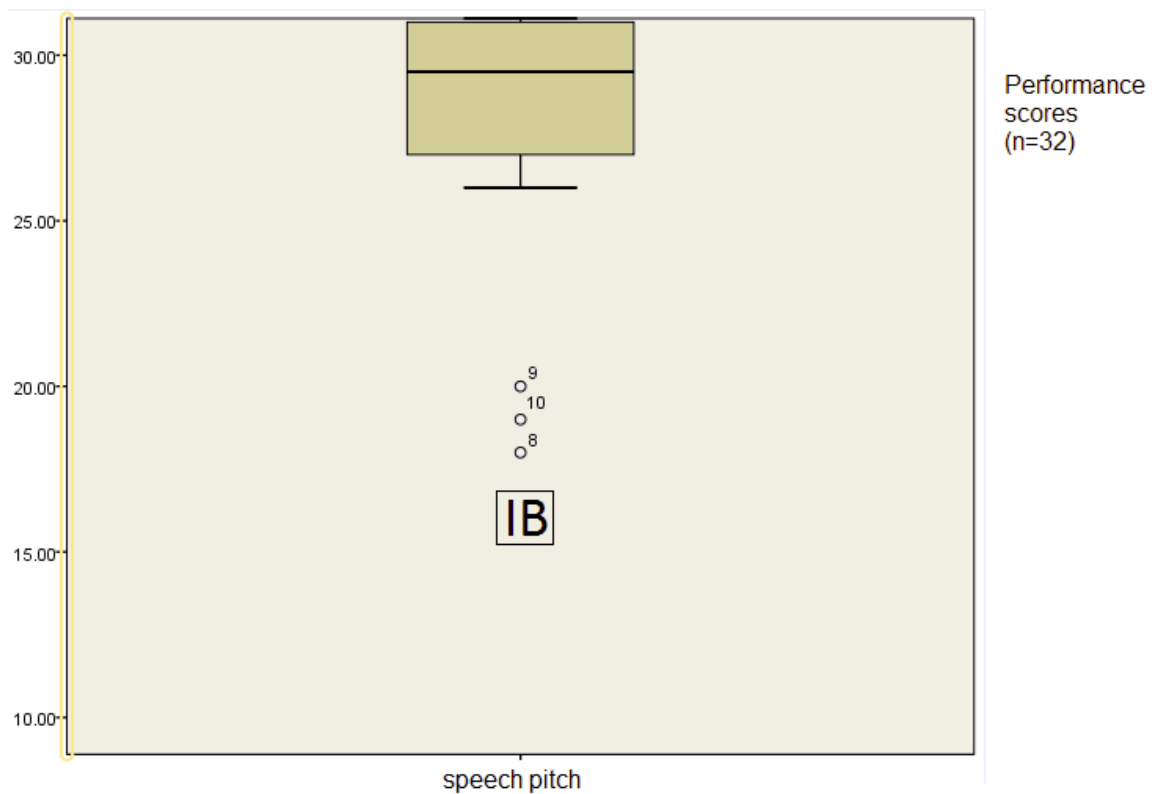
Figure 10: IB's performance on the speech pitch task relative to control participants. The horizontal line in the box shows the median. Participants number 9, 10, and 8 are outliers.

h) Speech rate. On the speech rate task, IB scored 16/32 (response mode: 3 possible options), whereas the control group scored 30.58/32 (min-max: 25-32, SD= 1.71). According to Crawford and Howell, (1998) significance test, the difference between IB and controls is statistically significant (t(29) = -8,354; p=0,000, estimated percentage of normal population falling below individuals' score = 0,00%). It is suggested that, as with the controls, it was easier for IB to compare linguistic rather than music stimuli in terms of speed, bearing in mind the duration differences across the two tasks.

g) Music tempo. On this task, IB scored near chance levels (10/32) and controls scored 23.29/32 (min-max: 17-29, SD= 4.15). Although not all controls performed at

ceiling, there is still a substantial difference in the performance of IB and controls. Even the weakest performance by a control participant (17/32) was substantially better than that of IB. According to Crawford and Howell, (1998) significance test, the difference between IB and controls is statistically significant (t(29) = -3,138; p=0,005, estimated percentage of normal population falling below individuals' score = 0,23%). This suggests that his ability to carry out this judgement was impaired. Figure 11 shows IB's performance relative to controls on the music tempo and the speech rate task.
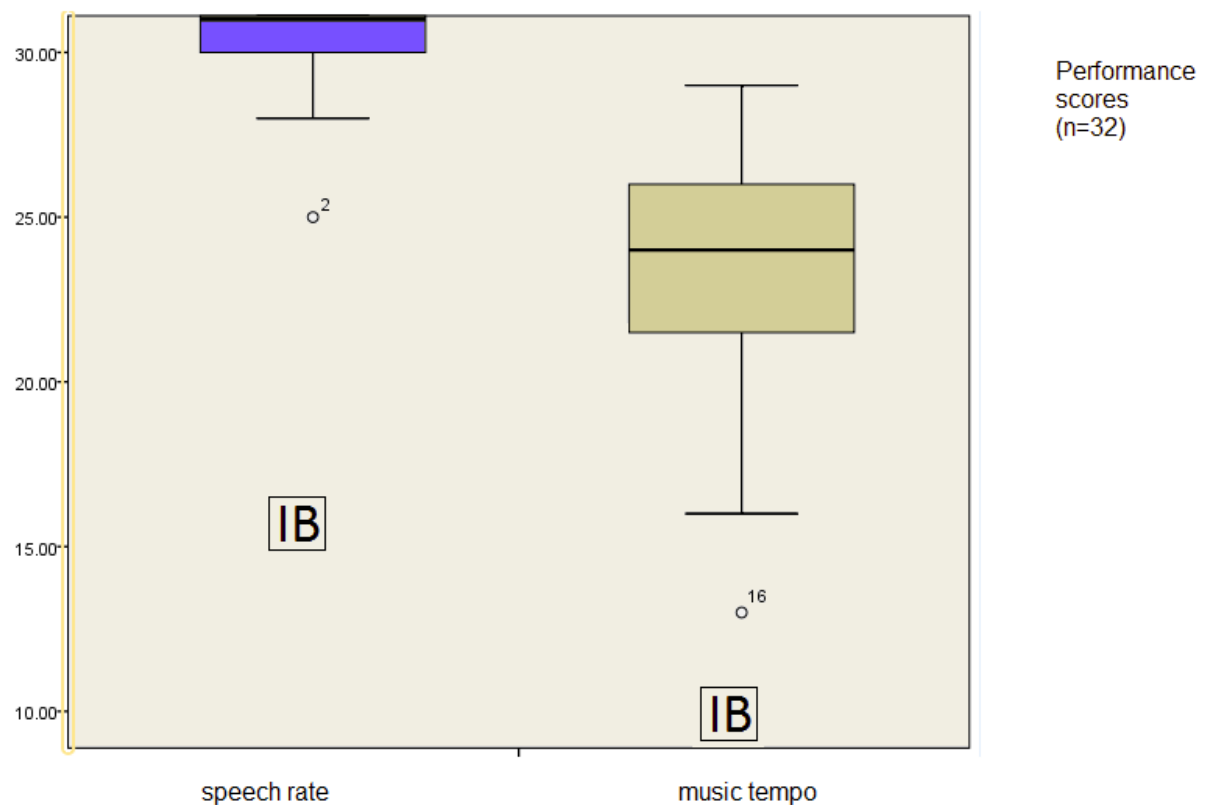

Figure 11: IB's performance relative to controls on the speech rate and the music tempo task. Dark horizontal lines indicate the median. Dark horizontal lines in boxes show the median in each task. Participant number 2 in the speech rate task and participant 16 in the music tempo task scored outside the range of the rest of controls.

In summary, IB's performance on these 8 tasks indicates that detection of speech prosody in a delexicalised context was severely impaired, in addition to discrimination between expressive and non-expressive speech prosody. Moreover, IB was impaired at perceiving pitch transpositions that took place in the second half of an utterance. Surprisingly enough, IB's ability to discriminate between expressive and non-expressive music prosody was preserved. Perception of emotion across domains was also intact in IB.

4.4 BZ's performance results

This section presents BZ's performance relative to neurotypical individuals on the 8 tasks of the LMP Battery. Results are accompanied by whisker plots that provide a visual representation of her performance.

a) Speech prosody detection. BZ's performance on this task was very high, achieving a score of 31/32. It should be noted that she outperformed some of the controls which was 28.83/32 (min-max: 19-32, SD= 4.12). These results suggest that BZ had no difficulty in identifying speech prosody in a low-pass filtered context. According to Crawford and Howell, (1998) significance test, the difference between BZ and controls is not statistically significant (t(29) = -0,516; p=0,611, estimated percentage of normal population falling below individuals' score = 69,46%). Her performance relative to controls is represented in Figure 12.
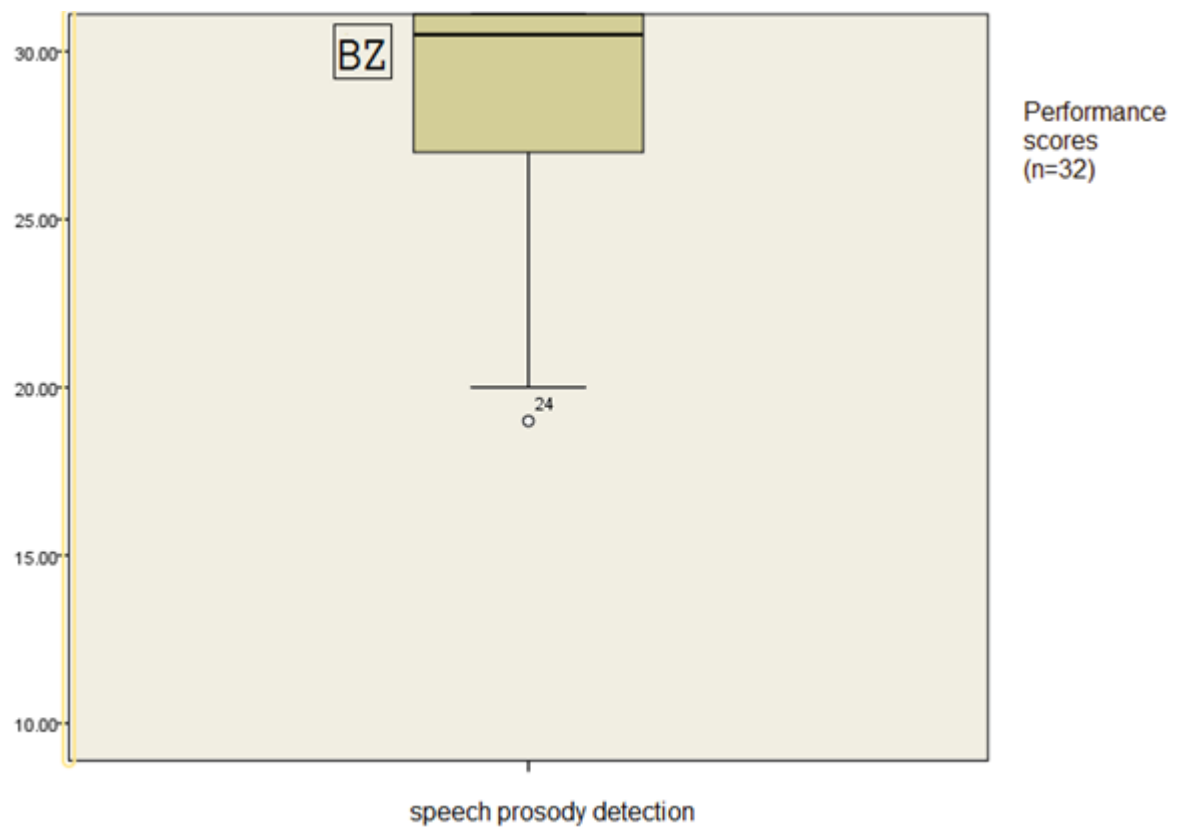
Figure 12: BZ's performance relative to controls on the speech prosody detection task. The horizontal line in the box shows the median. The circle shows an outlier, participant number 24.

b) Expressive music prosody. BZ scored very well on this task, despite her amusic profile. She achieved a score of 24/32 which was the mean of the control group, 24.16/32 (min-max: 17-29, SD= 3.15). There was a range of performance from the controls on this task, with one only scoring 17/32. According to Crawford and Howell, (1998) significance test, the difference between BZ and controls is not statistically significant (t(29) = -0,050; p=0,961, estimated percentage of normal population falling below individuals' score = 48,04%).

c) Expressive speech prosody. On this task, BZ scored 31/32, comparable to the
controls' mean of 30.16/32 (min-max: 25-32, SD= 1.49). According to Crawford and
Howell, (1998) significance test, the difference between BZ and controls is not
statistically significant (t(29) = -0,552; p=0,586, estimated percentage of normal
population falling below individuals' score = 70,70%).

BZ's performance on the expressive speech prosody task and its music counterpart
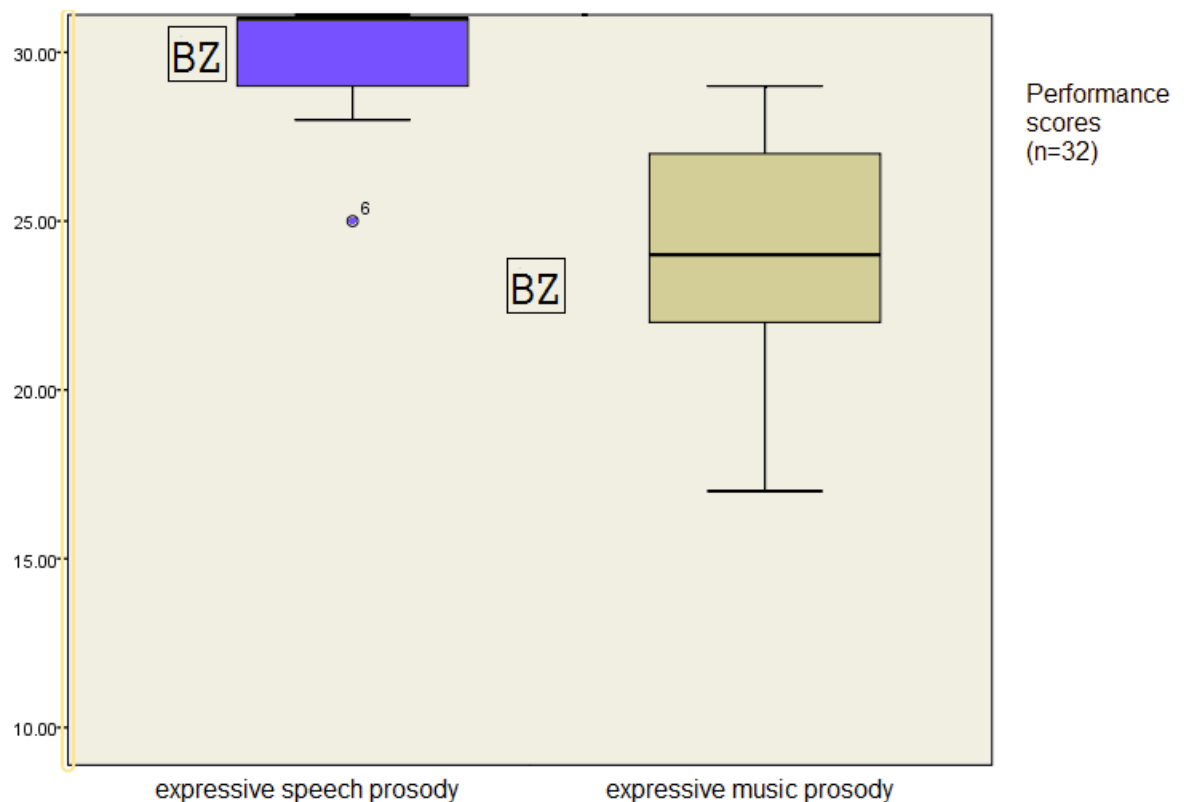relative to controls is presented in Figure 13.

**Figure 13: BZ's performance relative to controls on the expressive speech prosody and the expressive music
prosody tasks. Dark horizontal lines in the boxes show the median. In the expressive speech prosody task,
there is one outlier, participant number 6.**

d) Emotional speech prosody. BZ achieved the maximum possible score on this task; 32/32. The mean score of controls was 31.79/32 (min-max: 30-32, SD= 0.50). According to Crawford and Howell, (1998) significance test, the difference between BZ and controls is not statistically significant (t(29) = 0,412; p=0,685, estimated percentage of normal population falling below individuals' score = 65,77%).

e) Emotional music prosody. On this task, BZ was also successful with a score of 29/32, which was comparable to the control participants' mean score of 30.37/32 (min-max: 27-32, SD= 1.88). According to Crawford and Howell, (1998) significance test, the difference between BZ and controls is not statistically significant (t(29) = -0,714; p=0,482, estimated percentage of normal population falling below individuals' score = 24,12%). Figure 14 represents BZ's performance relative to controls on this task along with its speech counterpart.
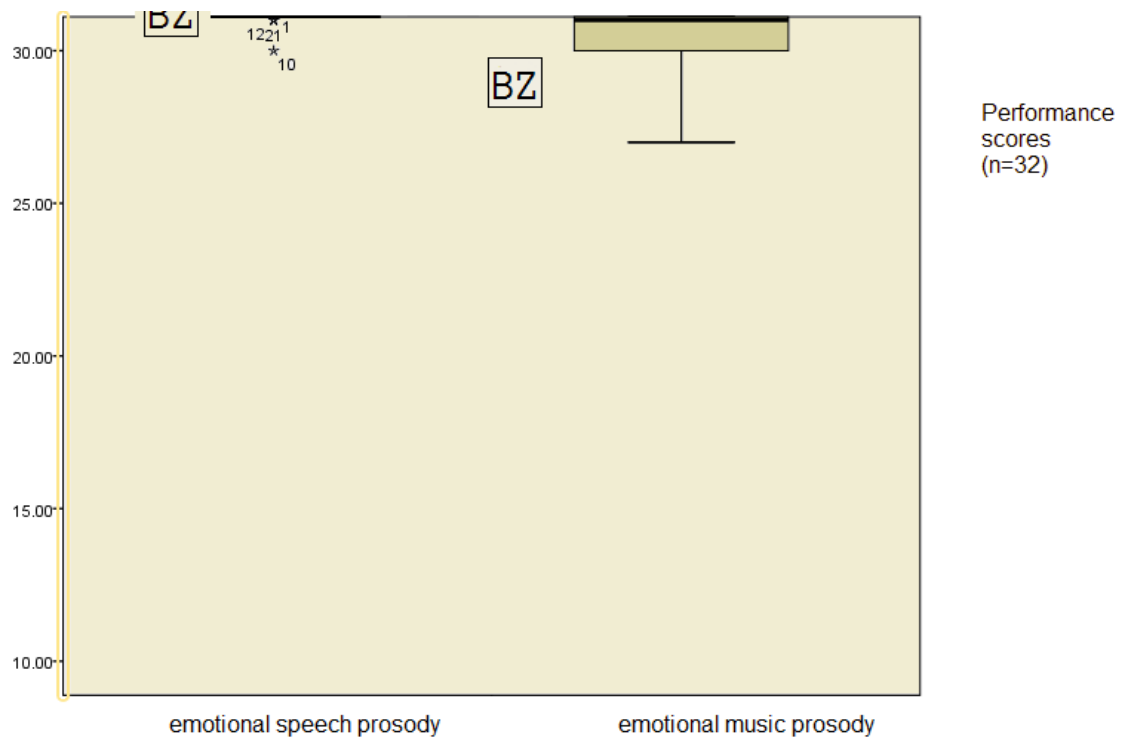
Figure 14: BZ's performance relative to controls on the emotional speech prosody and the emotional music prosody tasks. Dark lines show the median for each task .In the emotional speech prosody task, participants marked with a star (1,12,21,10) were the only participants that did not score 32/32. These participants scored 31/32 apart from participant number 10 who scored 30/32.

f) Speech pitch. On this task, BZ displayed her lowest score compared to controls. She scored 23/32, whereas controls scored 28.41/32 (min-max: 18-32, SD= 4.09). As her performance was above chance (recall that chance level for this task is 16/32), the present results suggest that she was not impaired at the speech pitch task. Rather, she might have had a problem with the format of the task for which, in comparison to IB, less detailed guidance was provided. As mentioned in 4.2, more than one third of control participants scored higher than 30/32 and it has been assumed that differences among participants that performed at ceiling and those that did not display a high performance were due to the format of the task rather than its degree of difficulty. According to Crawford and Howell, (1998) significance test, the difference between BZ and controls is not statistically significant (t(29) = -1,296; p=0,208, estimated percentage of normal population falling below individuals' score

156

= 10,39%). BZ's performance relative to controls on the speech pitch task is depicted in Figure 15.
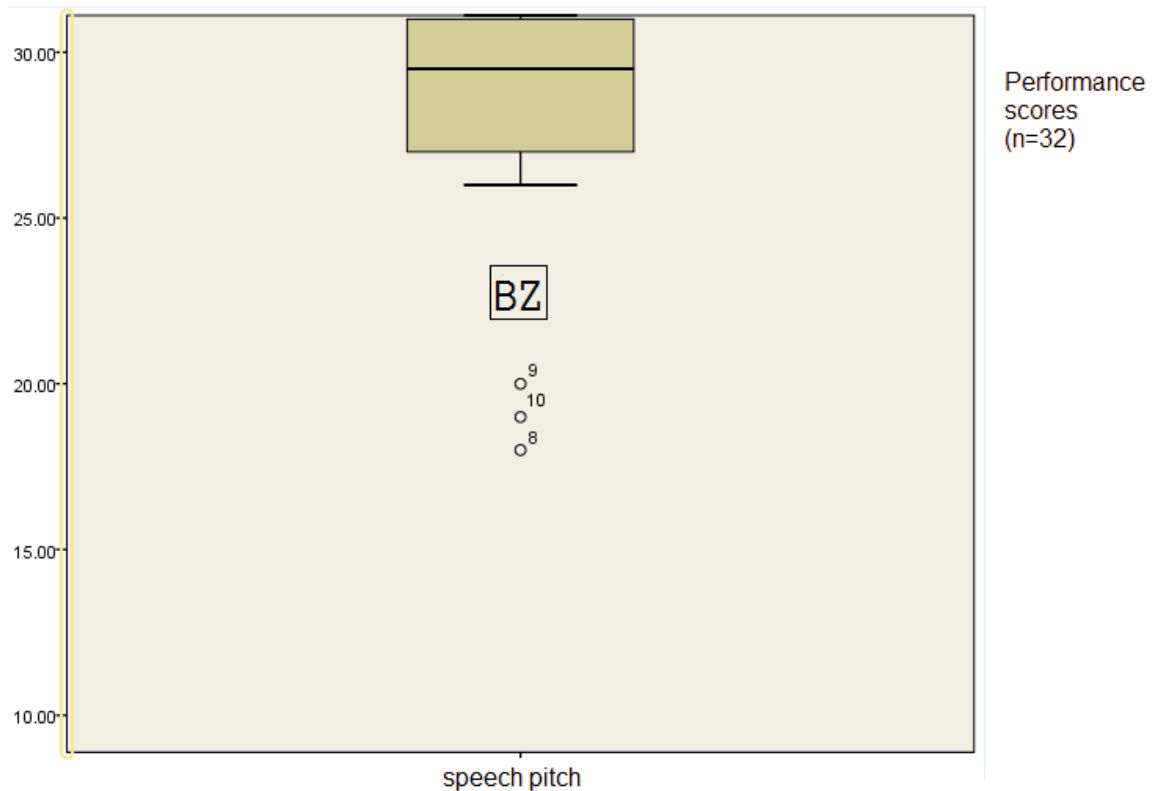
Figure 15: BZ's performance relative to control participants on the speech pitch task. The horizontal line in the box shows the median. Participants number 9, 10, and 8 are outliers.

h) Speech rate. BZ scored very high on this task, achieving a total of 31 out of 32 correct answers, comparable to the control mean which was 30.58/32 (min-max: 25-32, SD= 1.71) on this task. According to Crawford and Howell, (1998) significance test, the difference between BZ and controls is not statistically significant (t(29) = - 0,241; p=0,812, estimated percentage of normal population falling below individuals' score = 59,40%).

g) Music tempo. On this forced choice task including 3 answers, BZ scored 20/32. Her score was above chance (10.7/32). While it was lower than the mean control score of 23.29/32, her performance was still better than the lowest scoring control participants. The range of controls was: min-max: 17-29, SD= 4.15. According to Crawford and Howell, (1998) significance test, the difference between BZ and controls is not statistically significant (t(29) = -0,777; p=0,445, estimated percentage of normal population falling below individuals' score = 22,26%). BZ's performance on this task and its speech counterpart relative to controls is presented in Figure 16.
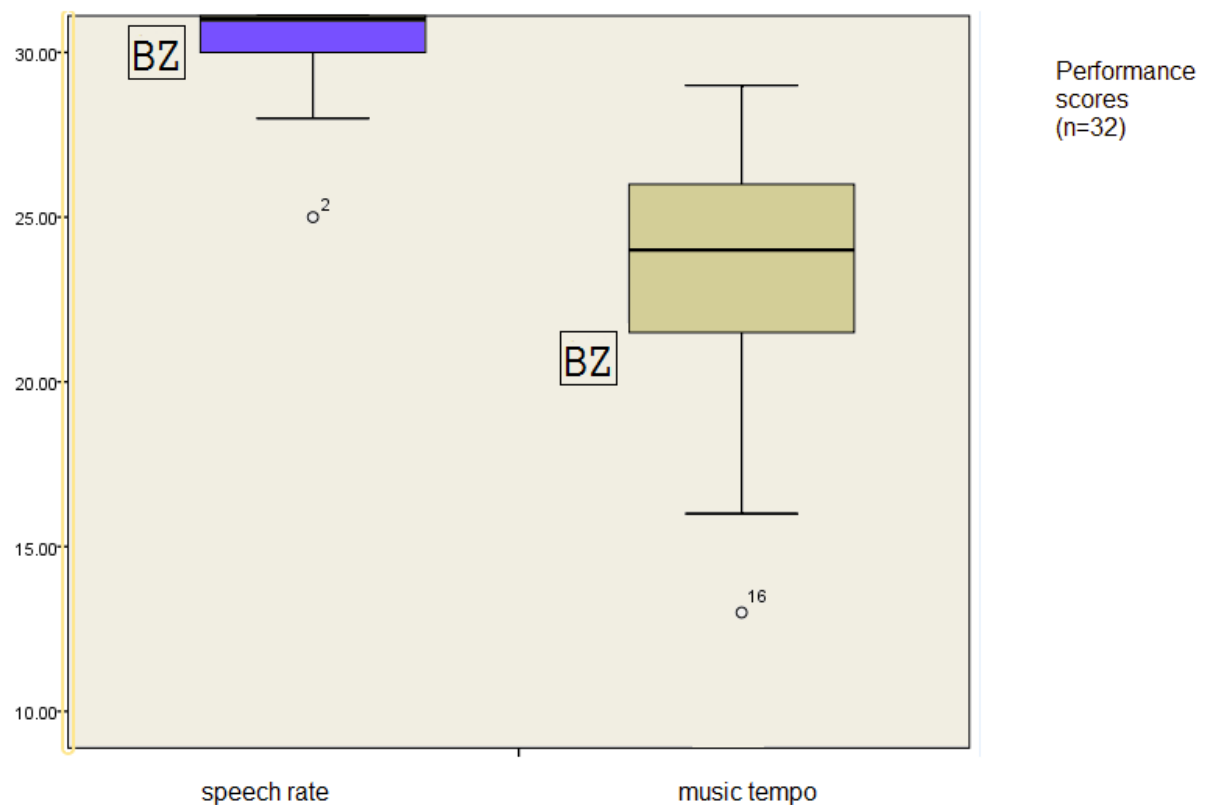
Figure 16: BZ's performance relative to control on the speech rate and the music tempo task. Dark horizontal lines in boxes show the median in each task. Participant number 2 in the speech rate task and participant 16 in the music tempo task scored outside the range of the rest of controls.

In short, BZ did not appear to be impaired at any of the processes that were tested through the above tasks. She performed at ceiling on the speech prosody detection

task, the expressive speech prosody task, the two tasks on emotion, and the speech rate task. Her score on the speech pitch task and the music tempo task was also within the range of controls. Despite being identified by previous tests as amusic, BZ was able to discriminate between stimuli in the novel expressive music prosody task.

4.5 Consideration of results from all participants

This section presents an overall picture of the results obtained from control participants, IB, and BZ. Although the controls were matched to IB and differed from BZ on age, education, and socioeconomic status, one can still appreciate some general points of comparison. The results in this section are presented for pairs of tasks in cases where for a speech task an analogous music task was also designed.

Table 3 shows the performance of control participants, IB, and BZ as well as the standard deviation and the performance range of controls. Table 4 shows the participants' performance in percentages as well as the percentages of guessing probability.

Tasks	Speech prosody detection	Expressive speech prosody	Expressive music prosody	Emotional speech prosody	Emotional music prosody	Speech pitch	Speech rate	Music tempo
IB	11/32	16/32	23/32	29/32	27/32	16/32	16/32	10/32
BZ	31/32	31/32	24/32	32/32	29/32	23/32	31/32	20/32
Mean	28.83/32	30.16/32	24.16/32	31.79/32	30.37/32	28.41/32	30.58/32	23.29/32
SD	4.12	1.49	3.15	0.50	1.88	4.09	1.71	4.15
Min-max	19-32	25-32	17-29	30-32	27-32	18-32	25-32	17-29

Table 3: Performance of controls, IB, and BZ. IB's and BZ's raw scores are provided. The mean of controls, their standard deviation as well as their minimum and maximum scores are also included in this table.

Task	Controls	IB	BZ	Guessing probability
Speech prosody detection	90%	34.3% ^	96.8%	25%
Expressive speech prosody	94.2%	**50% ^**	96.8%	50%
Expressive music prosody	75.5%	71.8%	75%	50%
Emotional speech prosody	99.3%	90.6%	100%	50%
Emotional music prosody	94.9%	83.4%	90.6%	50%
Speech pitch	88.7%	**50% ^**	71.8%	50%
Speech rate	95.5%	50%	96.8%	33.3%
Music tempo	72.7%	**31.2**%	62.5%	33.3%

Table 4: Performance of control participants (n=24), IB, and BZ reported as percentages on all tasks. Scores in bold indicate performance at chance. IB's performance marked with ^ indicates that IB favoured the same option in almost all his answers on these tasks.

a) Speech prosody detection. As shown in Table 3 the speech prosody detection task had a relatively large range of scoring. The lowest score of controls was 19/32. IB's performance was much poorer compared to the lower control score, while BZ was at ceiling. Although the probability of guessing was very low on this task (25%), given the four possible answers, it should be noted that IB had favoured the same option in almost all the questions of this task.

The whisker plot (Figure 17) adds to the analysis of the above data as it shows more clearly that the majority of controls had a good performance on this task.
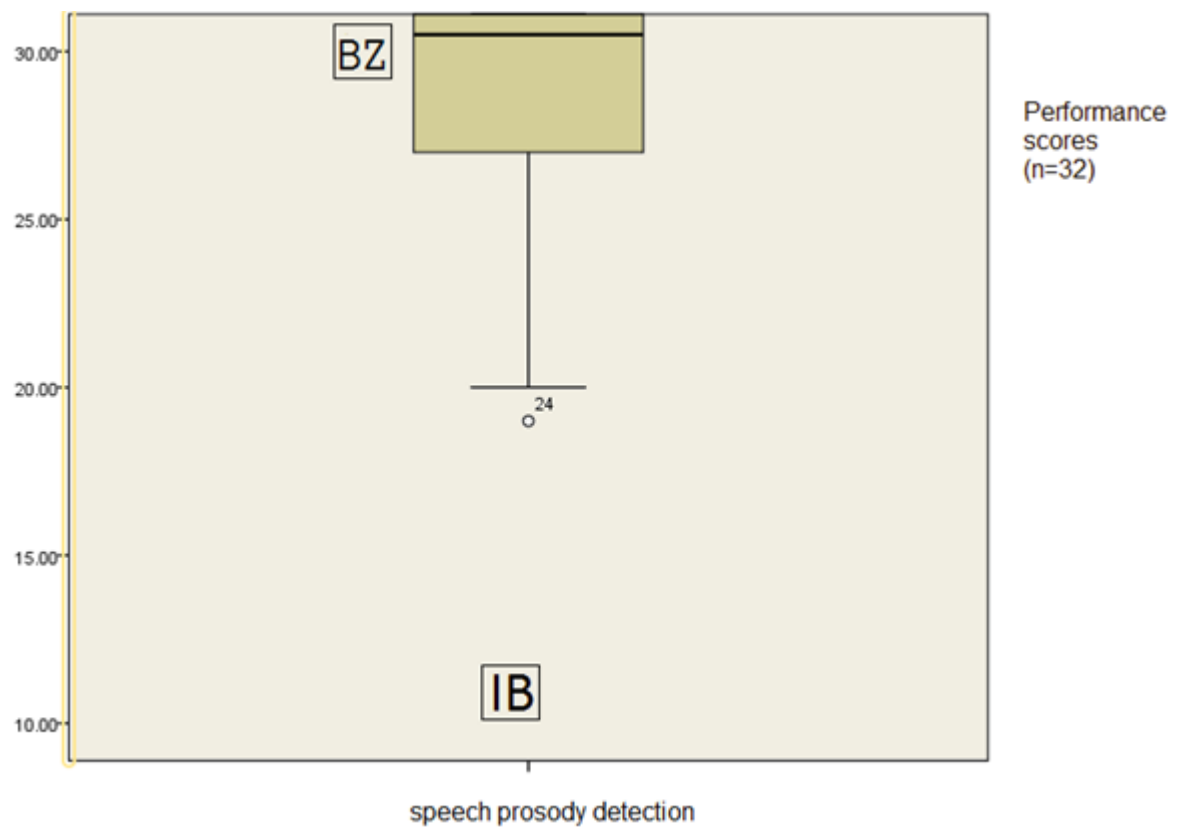
Figure 17: Performance of control participants, IB, and BZ on the speech prosody detection task. The horizontal line in the box shows the median. The circle shows an outlier, participant number 24.

b) Expressive speech prosody and expressive music prosody. As shown in Table 3 and Table 4, controls performed well on both tasks. However, they have a much higher mean score on the speech task compared to the analogous music task. BZ also performed similarly to controls with a relatively better performance on the expressive speech prosody task than the expressive music prosody task. The opposite pattern, compared to controls and BZ, was found in the case of IB; his performance on the expressive music task is much higher than the expressive speech task. The score of IB is similar to the mean performance of controls and BZ on the expressive music prosody task.

These patterns can be more easily visualised in the following whisker plot (Figure 18).
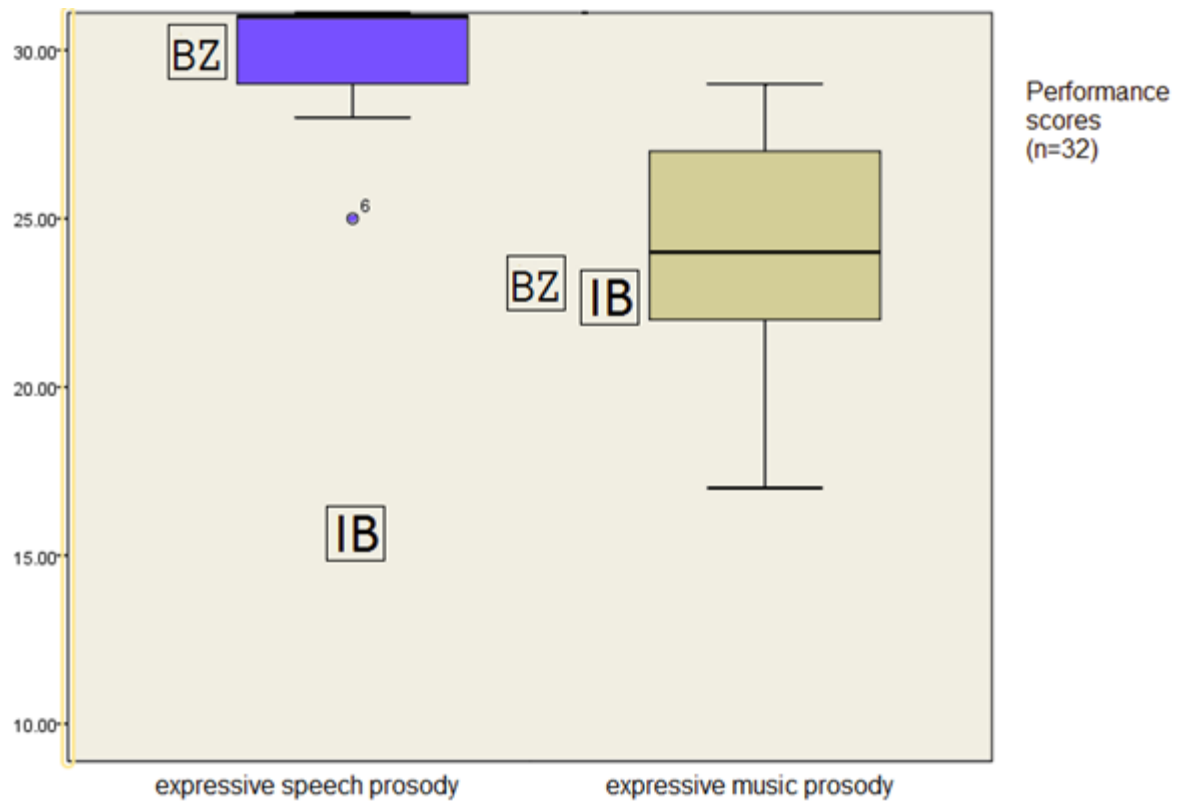
Figure 18: Performance of all participants on the expressive speech prosody and the expressive music prosody task. The horizontal lines in the boxes show the median. In the expressive speech prosody box, the median line almost overlaps with the upper line of the box. There is one outlier in the expressive speech prosody task, participant number 6.

c) **Emotional speech prosody and emotional music prosody.** For these two tasks, the performance of all participants, including IB and BZ was high. Performance of participants across these two tasks was relatively similar. As shown in Figure 19, all were able to perform these judgements very well.
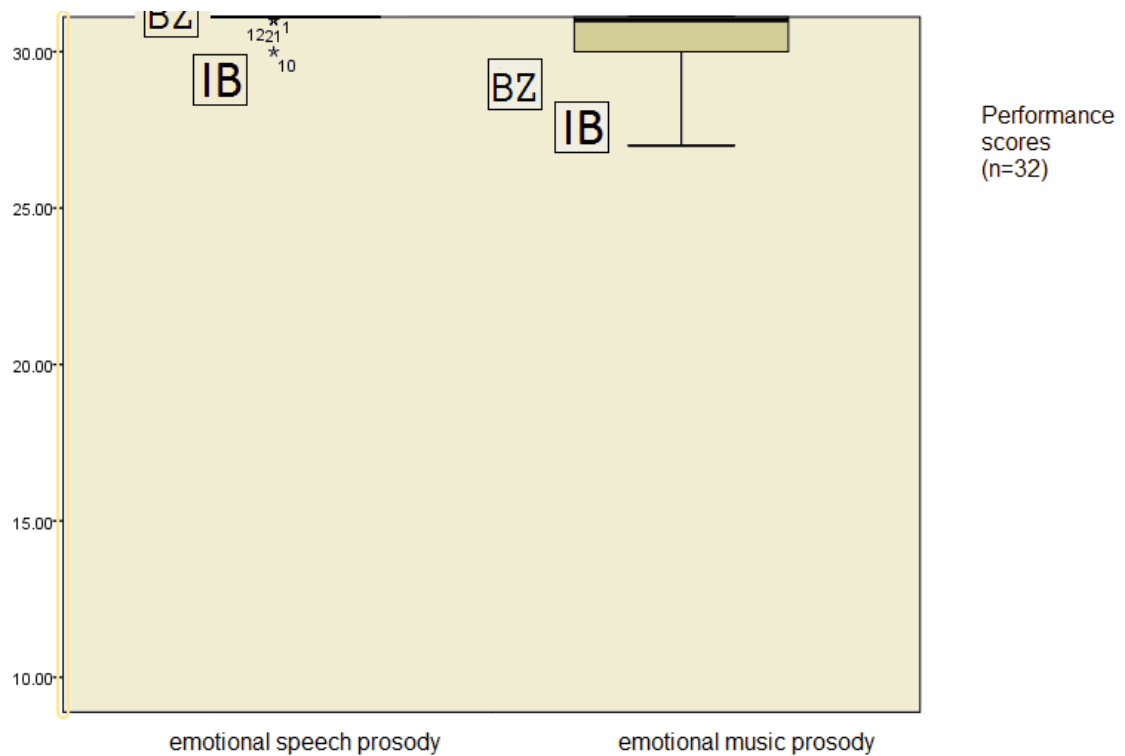
Figure 19: Performance of all participants on the emotional speech prosody and the emotional music prosody tasks. Dark lines show the median for each task .In the emotional speech prosody task, participants marked with a star (1,12,21,10) were the only participants that did not score 32/32. These participants scored 31/32 apart from participant number 10 who scored 30/32.

d) Speech pitch.

Most control participants achieved a good performance on this task. However, there was a fairly wide range of performance. The standard deviation for this task, as shown in Table 3, is also quite large. IB scored at chance on this task but BZ performed within the control range. The range of controls' performance is large in this task but, as it is shown on the following whisker plot in Figure 20, there were 3 particular individuals out of a group of 32 who performed particularly poorly. Nevertheless, IB's performance was even below the weakest performance by a control participant.
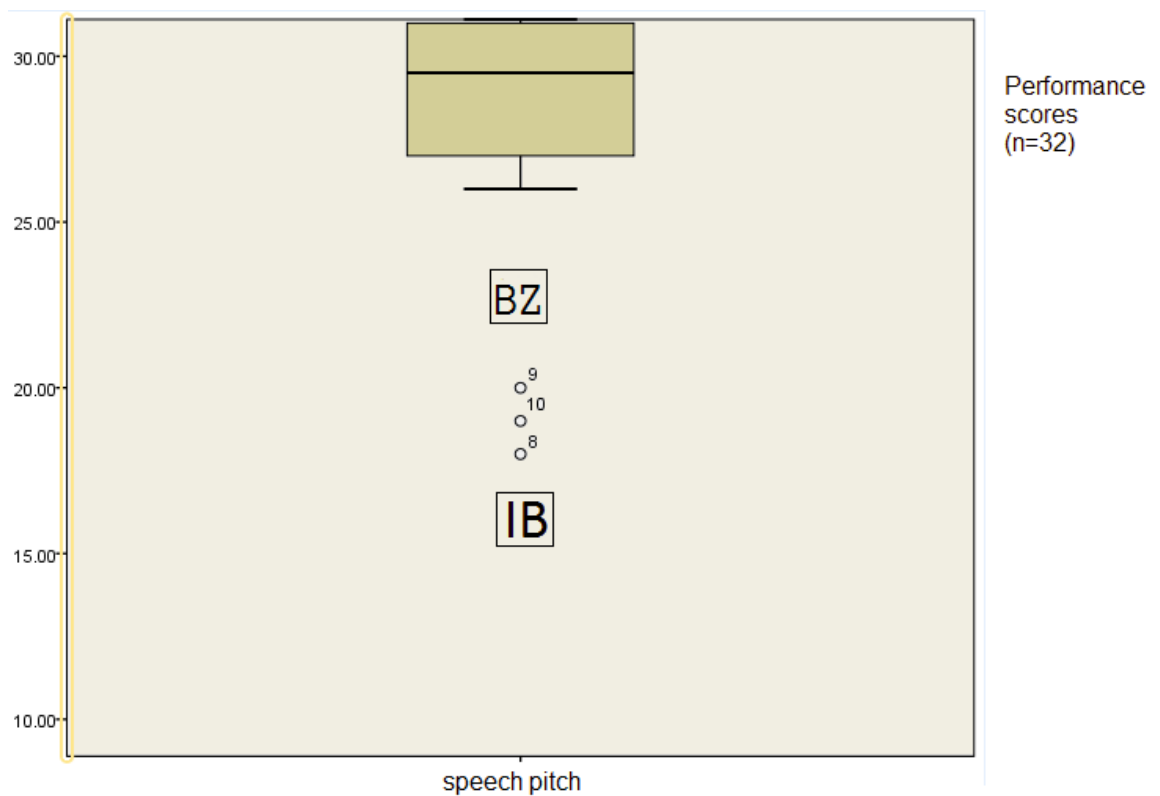
Figure 20: Performance of all participants on the speech pitch task. The horizontal line shows the median. Participants number 9, number 10, and number 8 achieved a score much lower compared to the rest of controls.

e) Speech rate and music tempo. These two last tasks show a ceiling performance in the speech domain and a much lower performance in the musical domain. The range of scores as reflected in the standard deviation is shown to be much larger for the music tempo task compared to the speech rate task in (see Table 4). The whisker plot below (Figure 21) shows that while there was a range of performance from the control participants, IB was impaired but BZ was not.
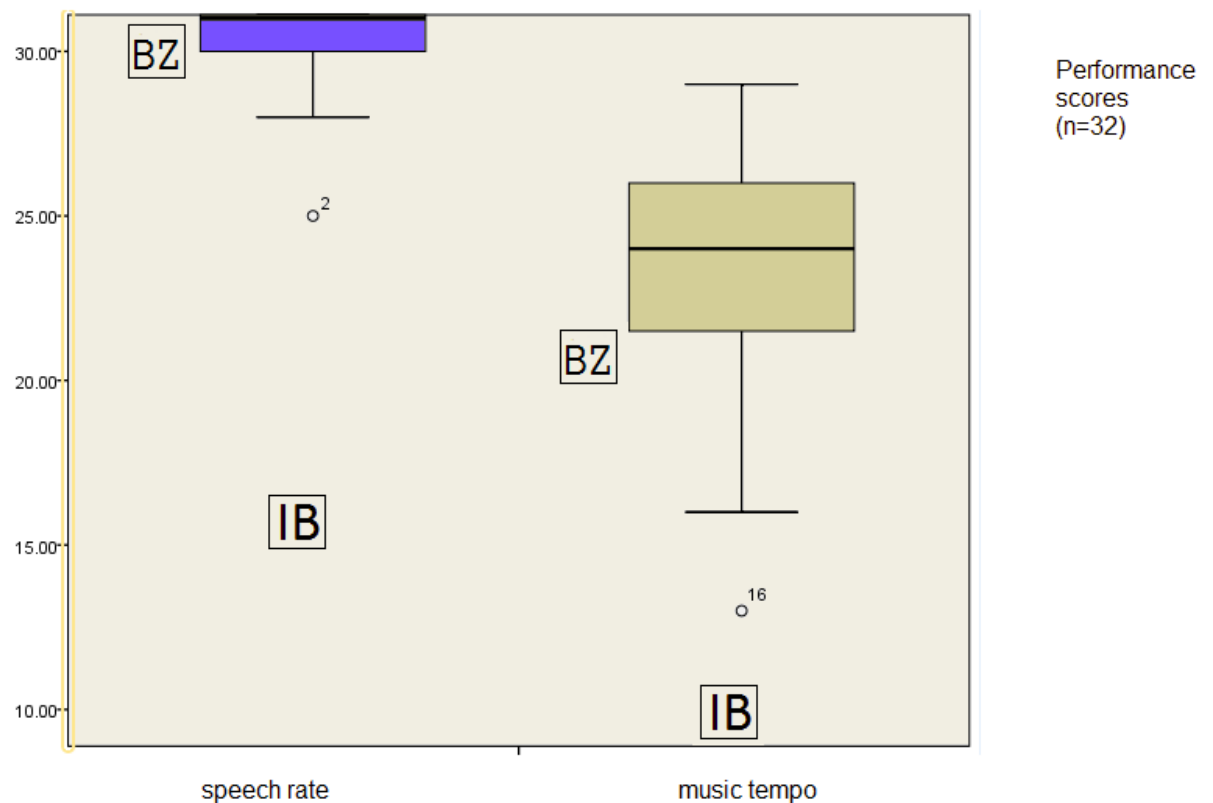
Figure 21: Performance of all participants on the speech rate and the music tempo task. Dark horizontal lines in boxes show the median in each task. Participant number 2 in the speech rate task and participant 16 in the music tempo task scored outside the range of the rest of controls.

In summary, taking into consideration the performance of all participants, emotion processing in language and music was an easy task for everyone. Expressive speech prosody was an easy task for controls and BZ, but this was not the case for IB whose performance was severely impaired. On the contrary, he performed normally on its music counterpart, namely the expressive music prosody task, as did BZ who was diagnosed as amusic in a previous study. The speech prosody detection task was a relatively easy task for all participants except for IB who favoured the same answer across most of the questions of the task. This tendency suggests that he was not able to identify any differences between the stimuli. The speech pitch was a task on which the average score of controls was high but a small number of them performed

less well, while still scoring above chance. These weaker performing controls were outperformed by BZ, whereas IB's performance was at chance level. Controls and BZ performed at ceiling on the speech rate task and IB scored above chance. By contrast he scored slightly below chance on its music counterpart, the music tempo task. Controls and BZ also scored lower compared to the speech rate task, but their performance was much higher than IB.

These results and their implications for the study of language and music as well as some methodological issues that arose are discussed in the following chapter.

5. Discussion

This study investigated prosodic language-music processing relationships extending previous research on pitch and emotion. These relationships were sought from an original perspective through the use of a novel Language and Music Processing Battery consisting of language and music tasks.

This section provides an initial treatment of the answers to the research questions presented in 2.7. Each of these questions is addressed explicitly here but a more detailed analysis of the answers in relation to previous literature is included in the following sections of the present chapter.

As stated previously, the most crucial question of the present study has to do with whether 'expressiveness' perception mechanisms are shared across the domains of language and music. Results from IB suggest that these mechanisms do not overlap, as his performance on the expressive music prosody task was similar to controls, whereas his score on the expressive speech prosody task was very low in contrast to

controls. Considering the results stemming from control participants, a different picture emerges, as their performance was much higher in the speech domain. Thus, results from both compromised and a neurotypical cognition suggest that processing 'expressiveness' is not equally demanding across domains.

The question on whether the present tools for assessing amusia are exhaustive methods of music cognition evaluation and the question regarding prosody perception in amusia can be answered based on the findings from BZ. Her performance on the expressive music prosody task shows that although an individual can be characterised as amusic, their performance on additional music tasks can be similar to the neurotypical listener. This calls for the design of additional tools aiming to define spared processed of music cognition in individuals that are diagnosed as amusics based on the MBEA. As far as speech prosody perception in amusia is concerned, BZ's performance on the speech prosody detection task and the expressive speech prosody task shows that speech prosody perception in amusics can be spared. More specifically, the task requiring identification of speech in low-pass filtered versions of speech with preserved prosodic properties and of speech with removed prosodic cues suggests that, in real life circumstances, an amusic individual can identify speech produced by a speaker in another room. Also, given BZ's performance on the expressive speech prosody task, an individual diagnosed as amusic can distinguish between utterances of different prosodic realisations.

Another research question of this study related to whether pitch differences of chunks of utterances (rather than individual syllables or tones) would be perceived by IB and BZ. IB's score on the speech pitch task revealed an impairment of this process, as his performance was at chance. This suggests that the voice quality difference that was produced as a result of the pitch manipulation that was applied

was not perceived by IB. BZ performed better and above chance but her performance was lower than that of controls. However, it is not clear whether her performance can be considered as compromised, given that she was successful on a large number of questions.

The last research question pertained to the possible differences between individuals with some musical education and others with no music education. This question, as it was previously highlighted, does not address all levels of music experience and is not perfectly in line with the distinction between 'musicians' and 'non-musicians' found in the literature, as number of the years of music training of the participants of this study is small. However, the results that were obtained in the study contribute to the understanding of the music listening abilities of non-musicians. More specifically, all participants performed similarly and no statistically significant differences were found between the performance of controls with musical training and those with no musical training. This finding suggests that the tasks that were designed for this study assessed musical abilities that appear to be universal and that do not depend on explicit music training. Individuals with musical experience did not perform better on the perception of 'expressiveness', emotion, pitch, or rate in comparison to participants with no music training. This study, therefore, contributed to the exploration of music cognition on the level of the unsophisticated listener.

The rest of this chapter offers a more comprehensive analysis of the above questions, adding a discussion of some methodological issues. More specifically, it discusses the effect of musical experience on linguistic and music processing according to the performance of control participants on the LMP tasks, some links between the performance of the right temporo-parietal brain-damaged male patient, IB, and previous studies, and some implications of the performance of BZ for the study of

congenital amusia. The present chapter also provides a treatment of the gap that the task variable of 'expressiveness' fills on the level of prosodic description, which is here argued to be the most important contribution of the study. Some general methodological implications arising from the investigation of pitch processing in this study are also analysed. Finally, the overall significance and contribution of this study to the parallel study of study of language and music are discussed, acknowledging limitations and possible confounding factors.

5.1 LMP Battery

The Language and Music Processing Battery (LMP) was designed for this study to assess the perception of certain acoustic features in the linguistic and the musical domain. This battery differs from previous tasks looking at the perception of speech and language as these used music-like stimuli derived from speech utterances rather than actual music stimuli. More importantly, it differs from previous studies as it included tasks designed to investigate a new type of speech and music prosody processing, here termed 'expressiveness'. In addition, the tasks of this study employed more naturalistic stimuli across domains than in previous research where the goal of matching analogous stimuli on timing appears to lead to distortion and a processing bias.

In order to appreciate the utility of the novel tasks designed for the LMP Battery it is more meaningful to look at the performance of the participants on its tasks. Performance on some tasks seems to bear important theoretical or methodological implications but other tasks appear not to raise similar issues. The following sections

refer only to tasks of meaningful implications according to the performance of either the group of controls or one of the two cases that were studied.

5.2 Controls: implications stemming from pitch processing and the excluded task

The tasks designed to test an aspect of pitch perception in healthy control participants raise a number of issues: the effect of identical manipulations of pitch in speech and music streams, the differences in perceiving short pitch events and longer ones, and the importance of taking into consideration methodological implications for the design of tools in the parallel study of language and music. The music counterpart of the 'speech pitch' task was only briefly presented in 4.2. As the performance of controls was very poor, it was thought that employing this task in the investigations of individuals with impaired processing would serve no meaningful purpose. However, the control data obtained from this task are worth consideration at the level of methodology in the parallel study of language and music.

The two pitch tasks required the detection of a pitch difference across a long chunk in speech and music pitch sequences. That is, in contrast to how pitch perception is traditionally tested, these tasks did not examine the perception of individual tones or the detection of off-key tones in a melodic context. Testing controls on these two tasks pointed to a pronounced dissociation of pitch perception across domains in healthy participants. The fact that control participants (both musically educated and musically uneducated) could complete the speech pitch task easily, whereas their performance was at chance level on its music counterpart is of interest. With respect to the analogous speech task, another thing that should be noted is that some of the participants reported listening to two different voices when they were exposed to the

stimuli. This suggests that there was a major perceptual difference between the first part and the second part of the spoken stimuli, when the pitch manipulation was applied. Such a pronounced difference in the musical stimuli does not seem to have occurred, as this manipulation did not result in a tone colour difference effect.

Hence, as many controls reported to have perceived parts of the stimuli on the speech pitch task as belonging to voices of different persons, it can be argued that this task did not test a pure pitch perception ability. That is, the pitch manipulation caused a voice quality difference which may not rely solely on pitch perception abilities. In other words, it can be argued that controls successfully discriminated between speech stimuli that carried a pitch manipulation extending over half of a phrase because they identified some difference in the quality of the voice rather than perceiving a pitch difference. This suggests that the same manipulation in the speech and the musical domain can, in fact, test different things. That is, in music, one will not necessarily perceive two music samples played by the same instrument as samples of different instruments if a small transposition has taken place. Put a different way, a melody starting two semitones higher than another melody played on the same instrument will still be perceived as played by the same instrument. Therefore, in order for the listener to detect the difference between two such melodies, they should be able to perceive that two melodies started on a different note. It is possible that placing the transposition in the middle of the melody results in an even higher degree of processing difficulty. The difficulty in perceiving such a difference in the music manipulation, while being easily perceived in the speech manipulation, demonstrated a strong dissociation across domains. This has important implications both at a theoretical level and at a practical level pertaining to the design of analogous tasks in the parallel study of language and music. The question

171

arises as to whether attention to detailed and accurate laboratory manipulations results in more reliable comparisons. This also pertains to whether precise manipulations contribute to the understanding of processing of naturally occurring linguistic and musical stimuli in everyday situations. It can be argued that if –for the sake of a 'safe' comparison– one manipulates stimuli to an analogous degree across domains, these stimuli might acquire a more similar nature than was present in the original naturally occurring sound streams. That is, manipulating music stimuli in order to make them comparable to speech stimuli might lead to them both being more likely to be processed in a 'speech mode', thus resulting into misleading conclusions about possible differences. Therefore, manipulations aiming to draw forth a very specific point of comparison can distort the nature of the original stimulus. After applying this type of manipulation across domains, claiming that instances of language and music are similar and are processed similarly might be biased by the 'accuracy' of manipulation. Alternatively, it can be suggested that it is the manipulated form of instances across domains and not elements of language and music themselves whose perception is proven to be similar or different. Hence, although technical manipulations can often guarantee more reliable testing conditions, examining isolated or manipulated acoustic cues that would not occur in a natural (out of the lab) context might not be legitimate.

At the same time, the conclusions drawn from these results appear to justify an assumption that guided the design of the expressive prosody tasks across the domains. As it was explained in 3.5, the rationale behind the creation of analogous stimuli was not based on applying exactly analogous manipulations across the domains. That is, applying identical manipulations across domains might lead to distorted versions of a speech or music stimulus that do not resemble naturally

172

occurring stimuli. Moreover, a given manipulation might have a completely different effect in the speech domain compared to the musical domain and vice versa. Controlling for a unique pitch manipulation and applying it across domains does not guarantee that the same type of processing will be engaged across domains. Finally, making two stimuli across domains comparable by using, for example, music-like versions of a speech stimulus instead of a real music stimulus does not seem to constitute a trustworthy method of studying language and music processing.

5.3 Controls: does music training always matter?

Although the effect of music training was not initially under the scope of this study, some of the findings that were obtained suggest that certain abilities in music processing do not exclusively depend on formal training but may also be developed through exposure to musical stimuli in everyday life.

Research investigating the effect of music on cognition has compared individuals with music training and those that have no exposure to musical training commonly referred to as non-musicians (e.g., Wong and Bradlow, 2005, Bidelman and Krishnan, 2010). Consideration of potential differences between musicians and non-musicians was deemed relevant to this study for two reasons. First, because previous research has demonstrated some differences in performing certain music perception tasks between these two groups. Secondly, the brain-damaged patient that participated in the study had musical experience. Therefore, a group of control participants with matched musical experience with the patient was also needed for the purposes of this study. At the same time, the existence of two different control groups, with and without musical experience, would allow for inter-control group

comparisons, in an attempt to determine whether musical experience was a variable affecting performance on the present novel tasks.

This study showed no significant difference on the performance of individuals with and without music experience on the tasks that were included in the LMP Battery. This may be due to the fact that the task demands were relatively slight and so did not call on what might be considered more specialised skills developed over time through practice. However, the perception of 'expressive' differences in musical performance (see 3.6.2) was one task which might have been expected to be performed less well by non-musicians. However, both groups were successful in this discrimination.

This may suggest the presence of some universal basic music abilities. While the productions of professional musicians can hugely vary, perceptual sensitivity to these features may be developed by simple exposure to music. That is, despite the significant difficulty in attaining the proficiency in producing these cues, the ability to perceive them only depends on listening to music, which often happens in a more passive way across the lifespan. In addition, being exposed to prosodically heightened speech, infants might develop this universal sensitivity early on. As discussed earlier, infants rely on prosodic cues at the first stages of language acquisition (e.g., Jusczyk and Krumhans, 1993, Snow and Balog, 2001). Hence, expressive speech and music can result in the development of this kind of aesthetic appreciation, regardless of formal training. It, therefore, appears that by listening to music across the lifespan, everyone has the potential to become an expert at differentiating between 'expressive' and 'non-expressive' performance. Although producing a performance with a rich expressive result is a laborious task for a soloist

or an actor, appreciating this effect in comparison to a deadpan performance seems to require no particular effort.

A word of caution is, however, necessary in interpreting these results. The 'musicians' that participated in this study were not professional musicians. Rather, they were chosen to have a music training profile of 3 years in order to match IB's music experience. It would be interesting to test professional musicians with decades of training and compare them to musically naïve individuals on similar tasks in order to determine if significant differences appear. However, as plasticity due to some types of training can induce differences over relatively brief periods of exposure (e.g., Herdener et al., 2010, Draganski et al., 2006), it was of interest to discover that individuals with three years of music experience did not perform better than musically uneducated controls on this task.

The results of this study suggested that the ability to perceive musical 'expressiveness' is a cognitive process that does not rely on training. As the two groups of participants, musically educated and musically uneducated individuals, were successful in performing these tasks, it seems that being causally exposed to music in the environment is sufficient for developing this perceptual sensitivity. As not all controls scored at ceiling, future research will possibly show whether additional training brings about a statistically significant difference between musicians and non-musicians. Given that this is a new variable to be studied in this comparative language-music research, experimenting with different music styles and improving this task in general is likely to provide a clearer picture.

5.4 IB

IB's performance on the LMP Battery demonstrates a number of spared abilities which were within the range of the healthy control participants while deficits appear to be 'task-specific' and selective. A generalised auditory discrimination difficulty was ruled out in baseline testing. A generalised cognitive difficulty can also be ruled out. More specifically, his poor performance on several tasks does not seem to be due to executive function and working memory demands in comparing acoustic stimuli in general. If this was the case, he would be probably equally impaired at all tasks requiring same-different discrimination judgements. His performance on the expressive music prosody task, for example, which was demanding for some controls, was high. His neurological profile is in line with this argument. Frontal areas that have been related to orchestration of executive processing and, more specifically, working memory and comparison judgement (Miller and Cohen, 2001), were intact in IB. IB's right frontal cortex was spared and this suggests that his compromised performance on a number of tasks can be attributed to his difficulty in perceiving the acoustic information of a given task, rather than to problems in information retention and comparison.

In other patients previously reported, retention and comparison of pitch and temporal patterns, but not extraction of information from the acoustic signal itself, have been attributed to the right frontal cortex (Patel et al., 1998). IB's good performance on the baseline stimuli on pitch in combination with his less good performance on the speech pitch task of LMP Battery point to dissociation of pitch processing between different time domains. The findings on the performance of IB on the speech pitch task seem to corroborate previous work addressing the differences of processing pitch in different time domains. That is, although IB's perception of pitch on this task was compromised, his perception between two pitch events in speech and music

stimuli was intact, according to the baseline tasks that he completed. This seems to offer additional support for the idea that an acoustic cue is processed differently in different time domains. As mentioned in 2.2, Poeppel (2003) formed the hypothesis that, although speech stimuli activate both hemispheres, the two hemispheres participate in a different way, as far as the temporal dimension of a given stimulus is concerned. More specifically, according to Poeppel, the areas of the left hemisphere pertain to short auditory signals, whereas right hemisphere regions subserve processing of longer acoustic streams. IB's performance is in line with the predictions arising from this hypothesis. That is, it can be argued that IB was successful in the comparison of short pitch events due to his intact left hemisphere, but his performance on a pitch task of a different time domain was severely affected by his lesioned right hemisphere.

Due to the original contribution of this study with respect to the element of 'expressiveness', it is not possible to compare IB to other patients' perception of 'expressiveness' as this has not been studied in previous neuropsychological literature. The study by Mazzoni et al. (1993) is the only investigation that, to the author's knowledge, makes a reference to a right temporo-parietal brain-damaged patient's global music sound appreciation. More interestingly, despite the similar lesion profile of IB and the patient in Mazzoni et al.'s (1993) study, the opposite performance pattern is observed. In their study, the self-report of a patient with right temporo-parietal brain damage reveals that the patient had lost his appreciation of the emotional content of music and also the ability to convert music perception into a meaningful experience. Based on this report, researchers suggested that the right hemisphere might be involved in functions of global sound appreciation. This is not in line with the findings in the case of IB, as his right temporo-parietal lesion did not

177

impair his ability to discriminate between 'mechanical performances' and performances with variation in dynamics, articulation, and deviation from temporal regularity. Put it another way, in contrast to Mazzoni et al.'s (1993) patient, IB was able to appreciate music streams globally in a way neurotypical individuals seem to do. However, these two sources of evidence should not be necessarily deemed conflicting, as in Mazzoni et al. (1993), their conclusion stemmed from the patient's self-report and not his performance on structured tasks as it was the case with IB.

An interesting observation arises from another speech task for which there was no music counterpart. In testing sessions for the purposes of the present study, all participants were first presented with the speech prosody detection task which included two types of stimuli. In this task, half of the stimuli included speech in which the only manipulation that was applied was that the lexical information was missing due to the removal of some frequencies in the acoustic signal. The remaining stimuli displayed no pitch variation and were also low-pass filtered. The combination of these two manipulations resulted in single tone pitch sequences which many participants characterised as 'Morse code' sound in the post-test debriefing. The listeners were asked to indicate whether the stimuli they were exposed to were speech or non-speech stimuli. In reality, all stimuli were derived from speech but the manipulations on one group of stimuli resulted in a non-speech effect. Control participants performed successfully on this task. IB, however, could not differentiate between the two groups of stimuli and sometimes commented on his inability to perceive any differences between stimuli. In fact, the majority of his answers favoured the answer 'both stimuli are the same', even in the cases that the stimuli belonged to different groups. This seems to be in line with the hypothesis that the left hemisphere processes the segmental features of speech and the right hemisphere

has a more critical role in processing prosodic features over larger chunks of a longer temporal span (Zatorre et al. 2002; Poeppel, 2003). As the stimuli were low-pass filtered and phonological information was missing, it can be argued that the preserved mechanisms of the left hemisphere did not have any facilitation effect on IB's performance. However, in contrast to evidence suggesting that subcortical structures participate in the processing of intonation contours, preserved subcortical structures in IB did not seem sufficient to support his ability to perform this task. Research with neurotypical participants suggests that the basal ganglia are primarily activated when one is exposed to intonation contours lacking lexico-semantic information as opposed to normal speech (Meyer et al., 2004). These structures despite being intact in IB did not lead to performance within the range of the healthy controls.

As far as emotional prosody is concerned, as this was assessed by the present LMP Battery, IB's performance seems to differ significantly from other right hemisphere-damaged patients. Patients with right brain damage (e.g., Gorelick and Ross, 1987, Ross and Monnot, 2008) and with right temporo-parietal damage specifically (e.g., Heilman et al., 1975) have been found to have impaired perception of emotional prosody. Although IB sustained damage in this area, his ability to judge emotional prosody was intact. This finding appears to run counter to prediction and highlights again the complex nature of studying prosodic impairments. However, it may be the case that IB did have some difficulty in emotional processing in the acute phase of his illness but this has since resolved. It is worth looking more closely at some studies that produced different results in terms of emotional prosody processing.

In Bowers et al. (1987), right hemisphere-damaged patients were impaired at processing and extracting emotional prosody. Controls and left hemisphere patients

were shown to have intact perception and categorisation of emotion in speech. Similar findings are reported in Blonder et al. (1991) where right hemisphere patients again, in contrast to controls and left hemisphere patients, were unable to perceive emotional prosody along with facial expressions and gestures of emotional content. More specifically, Heilman et al. (1975) tested right hemisphere patients on perception of four emotions; namely happy, sad, angry, and indifferent. Results showed that right hemisphere patients had compromised perception of emotional mood. The role of subcortical structures that were spared in IB might be the source of his intact ability. Subcortical structures, such as the right thalamus or the basal ganglia are thought to account –to some extent– for affective prosody processing, as bilateral cortical areas seem to interact with the basal ganglia (Kotz et al., 2003). Hence, these structures along with his intact left hemisphere might have minimised the hurdles posed by the compromised right hemisphere in processing emotional colour in speech.

With respect to perception of emotion in a musical context, the study by Baird et al. (2014) reveals some interesting methodological implications that appear relevant to the interpretation of IB's performance on the emotional music prosody task. In their study, a patient with a right temporo-parietal lesion, JM, displayed compromised recognition of sad and peaceful emotions in music stimuli, despite his high performance on general cognitive ability tasks. As IB, JM had musical experience as a drummer. Despite his impaired perception of sad and peaceful emotions in music, he had intact happy and scary identification ability. In particular, JM tended to identify sad music as peaceful. His musical cognitive profile shows impaired performance on the majority of the Montreal Battery of the Evaluation of Amusias subtests (contour, interval, memory, rhythm, scale) and average performance on only

one subtest, namely the meter subtest. That is, despite his generally compromised music cognition, he performed normally on the perception of meter. The authors suggest that selectively preserved meter is sufficient for the identification of happy and scary emotions in music but not sad and peaceful. This study shows that emotion perception does not depend on a single acoustic cue, at least, if one refers to a wide range of emotions portrayed in music. IB was able to perform at ceiling on the identification of sad and happy emotions, in contrast to JM that struggled with sad emotion identification. However, JM could have possibly performed better if he had been given a task including only sad and happy music stimuli that might be more easily distinguishable compared to sad and peaceful. It is also possible that IB might have been unable to identify happy and sad emotions successfully if these were embedded in a task with additional emotions such as peaceful and scary.

It, therefore, seems that it is important to take into consideration factors that might account for the differences in performance between IB and patients in other studies. IB was tested on the recognition of only two emotions and these results are not perfectly comparable to other studies, as emotional prosody processing was not the main scope of this study. The examination of his emotional processing abilities was not exhaustive, as the stimuli were relatively easy and more demanding stimuli might have pointed to some difficulty. A more general explanation on variability in performance on emotional prosody tasks across studies is the fact that, despite the use of common prosodic labels, different researchers employ different tools that possibly capture different prosodic aspects of the same label. That is, a patient might display emotional processing deficits due to pitch perception deficits whereas another patient might also have emotional prosody processing difficulties due to speech rate or loudness factors. As investigations of emotional prosody might be

designed with more emphasis on different acoustic features, performance of patients in different studies cannot be always compared objectively. A patient, for example, might be impaired at perceiving speech rate which might affect his perception of sadness versus happiness. Another patient, displaying an impairment in the perception of voice quality, might also display an impairment in distinguishing sentences with happy and sad emotional tones. As the design of the stimuli might have produced the difference between emotional prosody aspects focusing on different acoustic features (e.g., speech rate, voice quality, pitch variation) patients across different studies might be shown to have emotional prosody impairments subsequent to different types of deficits.

The dissociation that IB displayed between expressive speech prosody and expressive music prosody is treated separately in 5.6, as it relates to the concept of 'expressiveness' which is central to the present study.

In this short section, some tangential links to previous literature were presented. Future investigation using the same tasks with other patients might cast some light on the nature of IB's impairments and the mechanisms involved in the successful completion of the tasks designed for the purposes of the present study.

5.5 BZ

Compromised abilities of congenitally amusic individuals have been traditionally related to pitch or rhythm deficits. The Montreal Battery of Evaluation of Amusias (MBEA) (see 3.7.3) assesses a number of musical skills and depending on the acquired score can diagnose an individual as amusic or not. BZ who participated in this study had been tested previously on the Greek version of the Evaluation of

Amusias (GBEA) in the study of Paraskevopoulos et al. (2010). What is interesting about BZ is that, although having been characterised as 'amusical' by her performance on the Amusia battery, she scored normally on the expressive music prosody task of the present study.

BZ's performance on the GBEA and the expressive music prosody task of the present Battery suggests a paradox. An individual can be identified as having a developmental disorder of music processing despite being able to appreciate aesthetic features of musical performance which can be viewed as an integral part of music appreciation. This observation might call for additional subcategorizations of amusic individuals, as there are some individuals that find music disturbing and others that are also viewed as amusic and think that they are completely deprived of musical ability, but possess the ability to distinguish between performances and appreciate expressive versus deadpan versions of the same melody. This suggests that the term 'amusia' should be revisited. The necessity of further dividing this impairment has been acknowledged, as amusia has been shown to have a melodic and a rhythmic component (e.g., Peretz et al., 1990, Peretz et al., 1994). However, it can be argued that the term remains generic, as the subtests of the MBEA do not capture all aspects of musical ability. Shifting away from traditional ideas on musical ability might prove helpful in the understanding of music cognition and relationship with speech.

The case of BZ provides illuminating findings for the issue of speech prosody processing in congenital amusia. As it was discussed in 1.3, congenital amusia has been associated to speech prosody impairments but not a clear picture of this association has been hitherto presented. BZ scored at ceiling both on the expressive speech prosody task and the speech detection task. In contrast to previous studies

with amusics, BZ was not required to make discrimination judgements between statements and questions, stimuli with different grouping, or stimuli with differences on the words bearing emphasis in a sentence. She was required to make discrimination judgements between utterances with rich prosodic inflection and absence of prosodic inflection. BZ was successful on the task. At the same time, she displayed ceiling performance when asked to detect speech in a delexicalised context. That is, she was able to identify speech based on the existence of prosodic inflections. This shows that she was able to process speech prosody regardless of lexico-semantic content.

According to the findings obtained from BZ, congenital amusia can co-exist with spared speech prosody and 'expressiveness' perceptual abilities based on the specific tasks that were here employed. It is suggested that further progress in understanding the nature of congenital amusia might be made by researchers employing a combination of MBEA and LMP Battery tasks that could provide a fuller picture of music ability processing.

5.6 'Expressiveness' and performance

Perception of 'expressiveness', as assessed through the expressive speech prosody task and the expressive music prosody task, is here treated separately because it is argued to be the one of the main focuses of this investigation. Both the development of the theoretical concept and the novel tasks that were designed to gather objective evidence concerning it constitute original contributions to the parallel study of language and music.

It should be recalled that on the expressive music prosody task, performance was similar in the 14 musically trained control participants, 10 non-musicians control participants, 1 right temporo-parietal brain-damaged patient and 1 congenital amusic individual. None of the participants had any difficulty in discriminating between melodies with variation in loudness, articulation, expressive pausing, and deviation from temporal regularity from melodies that did not display these features. This suggests that the ability to make such a perceptual judgement may not vary according to musical training and can remain intact in the presence of extensive right hemisphere damage or congenital amusia. On the expressive speech prosody task, controls and BZ had a similar performance, but IB's performance was very poor.

Making judgements of 'expressiveness' in speech and music appears to pertain to different mechanisms, given the performance of IB. Both the speech and the music task assessed the ability to distinguish among different performance styles in speech and music. Despite the fact that both tasks had this element in common, IB's performance showed that making same-different discrimination judgements of expressive performance across domains relates to different perceptual abilities. This dissociation is also supported by the performance of control participants, who displayed the opposite performance pattern; they did much better on the speech task compared to the music task. This can be probably attributed to the easier processing of the speech stimulus given the presence of lexical information, despite the fact that it was semantically neutral. For controls, 'expressiveness' was more easily processed in the 'speech mode'. It is, however, important to note a possibly confounding factor that might –to some extent– account for this performance dissociation that was detected. Controls might have relied on the pitch variation difference between expressive and non-expressive stimuli in speech in order to achieve this ceiling

performance. This difference was not present in the music stimuli, as both expressive and non-expressive music stimuli were comprised of exactly the same pitch sequences. Therefore, pitch variation was not a differentiating factor in the music domain. At the same time, given that IB was impaired at detecting speech prosody contours (as the speech prosody detection task shows), his inability to perform well on the speech task might closely relate to his impaired speech contour perception. However, if this was indeed a confounding factor related to this dissociation, it is still very interesting that IB's ability to differentiate between expressive and non-expressive music stimuli was intact in the presence of extensive right hemisphere damage.

It is not known whether pitch variation perception had a facilitation effect for BZ. According to the results of the Amusia Battery on which she was tested by Paraskevopoulos et al. (2010), BZ's pitch perception in music was impaired. However, the tasks employed by Paraskevopoulos et al. (2010) included assessment of pitch in a tonal context. The tasks in the LMP Battery did not test pitch in this context and also pitch perception was embedded with other acoustic cues in some tasks.

With respect to the relationship between 'expressiveness' in language and music, BZ's performance was very similar to that of controls. This provides further support for the difference in the processing involved in the expressive speech prosody task and the expressive music prosody task. That is, although the two processes were spared in BZ, her performance was much higher in the speech domain compared to the music domain. However, since this pattern was also demonstrated with the neurotypical participants there is some question about the relative difficulty of the two tasks which requires further investigation. Hence, BZ's lower performance in

music 'expressiveness' compared to speech may not simply be attributed to her developmental deficit. Put a different way, perception of 'expressiveness' across domains seems to be preserved in the presence of congenital amusia but to different degrees, according to the data stemming from BZ.

Returning to IB's performance, one can argue that his former music experience might be responsible for his spared ability to distinguish between music performances. If this is the case, it is still interesting that this ability did not transfer to the speech domain. That is, if IB was successful on the expressive music prosody task because of his pre-stroke music experience, it is still interesting that this experience did not facilitate processing of 'expressiveness' in the speech domain. This again could suggest that the two tasks did not require the same type of perceptual judgement and this is why training in one domain might not have facilitated performance in the other.

It can be argued that 'expressiveness' does constitute a prosodic aspect in both the linguistic and the musical domain, but the dynamic combination of features that contribute to this aspect are likely to differ across domains. Generally, in speech, pitch variation seems to play a more important role in displaying differences between speech utterances. The pitch variation in this speech pitch task was possibly a useful discrimination cue between expressive and non-expressive stimuli. In piano music, different performances of the same melody do not include pitch differences and, therefore, appreciating differences in expression does not depend on pitch variations. Therefore, the stimuli of the analogous music task differed in terms of other acoustic cues and not pitch variation. What has not been answered by the present data is to what extent perception of pitch variation affected performance on the expressive

speech prosody tasks and what implications this might have for the comparison between 'expressiveness' in speech and music.

5.7 Do the established categories suffice?

The term 'expressiveness' was coined in this study in order to capture a novel prosodic aspect of the linguistic and the musical domain. Research on the prosodic realisation of storytelling also seems to raise questions about the sufficiency of linguistic and emotional features to provide an exhaustive description of speech prosody features. The prosody of storytelling and the concept of 'expressiveness' capture the necessity for the establishment of new prosodic categories that can account for the unique elements present in a sample of speech output rather than another.

5.7.1 Parallel developments suggesting future development: storytelling prosody

Converging evidence from a different research domain seems to confirm that suprasegmental elements other than linguistic and emotional prosody are present in the acoustic stream. Evidence on listeners' judgements of synthesised storytelling speech renditions pertains to another field of research in which additional aspects of prosody are acknowledged.

A new field of research in computer speech synthesis has developed in parallel to that of neurolinguistics. Some appreciation of the significance of prosodic elements has also become relevant in this distantly related domain of inquiry. Researchers working on speech re-synthesis for digital storytelling applications producing

computer-generated speech also seem to suggest that elements beyond emotion can have a significant effect on speech which, in their case, are represented by the label 'storytelling prosody'. Although the first storytelling applications attempted to reflect the expression of emotions, recent work pointed the need for an 'expressive' narrating style. In Callaway and Lester (2002) prose generation techniques led to many different retellings of a single story. This adds support to the argument that one cannot predict the acoustic realisation of a written text based on the information that it is provided in it. It also highlights the potential high degree of variability that can occur even in the absence of syntactical or emotional differences.

If one assumed that syntactic structure and expression of emotions would suffice to dictate a specific prosodic realisation of a sentence, then one would be able to consistently predict the prosody qualities of a spoken rendition of a written sentence. However, this does not seem to be the case. It has been accepted that there is a great deal of between or within-speaker variability (Aylett and Turk, 2004).

It is worth looking at the method employed by a study in order to capture some prosodic rules that seem to be typical in storytelling. Theune et al. (2006) generated a type of storytelling speaking style for children applications based on the actual speech of a human storyteller whose speech they analysed. Subsequent to the analysis, they designed and implemented a set of prosodic rules in order to transform speech with 'neutral' prosody into speech with storytelling prosody. This strategy seems to take into account the fact that speaker variance is not captured in simple and predicable phonological rules that could be fully predicted by syntactic structure or portrayal of specific emotions according to the content of the story. It is also worth turning to a point of comparison that the above researchers made. Comparing storytellers and newsreaders, they detected a number of prosodic differences in their

output. They found that in comparison to newsreaders, storytellers employed more variation in pitch and intensity. They also found that the storytellers spoke slower than the newsreaders and also tended to take longer pauses, especially among sentences. In fact, their pauses were even longer than one could predict based on their global differences in tempo compared to newsreaders. In addition, they added more emphasis on adjectives and adverbs through increased pitch differences and duration. In general, they used larger pitch excursions than the newsreaders. With respect to the judgements on prosody by those exposed to the speakers, the speech with 'storytelling prosody' was preferred over stimuli with neutral prosody. Although their study was solely about artificial computer-generated speech manipulations, an interesting point can be raised based on their findings with respect to the study of prosodic elements in language.

The storytellers were found by Theune et al. (2006) to place more emphasis on adjectives and adverbs which may point to a rationale for the development of novel tasks to investigate expressive prosody in speech. However, this is also of limited value, as such an observation could not be extended to 'expressiveness' in music. This reflects important differences between music and speech with regard to their formal structure. While speech prosody may be organised suprasegmentally around semantic and syntactic structures, music does not have parallel properties in this respect. As music does not contain parts of speech such as adjectives or nouns, placing emphasis on a musical phrase will be guided by other principles. Hence, despite the fact that manipulating features such as duration, pitch, or loudness appear to constitute expressive means for both language and music, their use depends on distinct structures across domains.

It is important to narrow-down the argument that speech prosody cannot be predicted on a written text to the type of spoken utterances one refers to. For example, when it comes to news broadcasters, the degree of unpredictability might not be particularly high. In contrast, when it comes to storytelling, variation appears to be much larger. Doukhan et al. (2011) recorded and analysed a one-hour long corpus of stories by a French speaker. They compared storytelling style with political address and radio news speaking style in terms of prosodic variation. Their findings pointed to a larger prosodic variation for storytelling. 'Expressiveness' (referred to as 'expressivity' in their research) is a result of changes in pitch and intensity registers, use of glide tones, and devoicing which were found to be employed for the differentiation of structural segments of tales. With respect to pitch range, they found a 6 semitone difference between storytelling and the other speaking styles they looked at. The beginning of the tales displayed high glide rate and inter-syllable pitch difference and the epilogue combined low pitch, low intensity as well as low speech rate.   In general, storytelling displayed more prosodic variations compared to other speaking styles, being of significant usefulness to speech synthesisers, as its features enhance expressivity. This study also showed that expressivity moves beyond expression of emotions and syntactic structure. If emotion had a central role in expression, tales would be probably organised around emotional tone. Moreover, if syntax governed organisation of prosodic structure, no differences would have been observed among storytelling speaking style, political speaking style, and news speaking style. These observations highlight a significant difference between language and music on the level of 'expressiveness'. Different registers such as political speech and radio news do not have counterparts in music, as these sociolinguistic aspects cannot be easily

found in the music domain. Although register in the sense of style is part of a music tradition, it does not bear the exact characteristics of language in its social context.

The fact that a story can be retold in different styles and also the differences in prosodic choices of storytelling speech and other speech styles point to a very promising candidate for additional study of prosody in general. Studying prosody through storytelling but also through the observation of acting performance can shed light on the richness of unexplored aspects of prosody. The study of storytelling prosody as discussed here appears to be in line with the investigation of 'expressiveness'. Large intra- and inter- speaker variability calls for thorough investigation of additional aspects beyond syntax and emotional prosody. The additional dimensions that emerge might be higher to grasp and define but their existence is proven through all the cases of variability that have been so far identified. Moreover, if prosodic cues in speech and music are governed by other factors that are domain-specific, they themselves might be deemed as domain-specific, in spite of the use of common acoustic cues serving their realisation. These observations can possibly pave the way for the establishment of a new type of detailed prosodic description, based on a larger corpus of data in the future.

5.7.2 'Expressiveness': Capturing an additional prosodic dimension

Research on generation of storytelling applications tangentially relates to 'expressiveness' in the linguistic domain. Storytelling application literature, as it was discussed, reveals the researchers' concern for more natural prosody in oral storytelling applications that extend beyond the portrayal of emotions and syntactic relationships.

The main research question of the present study was shaped by reflecting on how one appreciates the aesthetic value of a linguistic or a musical performance. As in speech, music has prosodic cues that contribute to an aesthetically interesting result that can lead to the appreciation of one performance rather than another. It was, therefore, attempted to examine whether the compromised cognitive system 'recognises' differences between expressive and non-expressive performances and if this differs across the domains of language and music. The concept of 'expressiveness' has been identified here from an exclusionary perspective, that is, by identifying relevant gaps in the literature that had previously investigated prosody. It was demonstrated that not all prosodic cues are encompassed by the existing prosodic labels. In a more theoretical perspective, it was noted that this new prosodic aspect cannot necessarily been broken down to specific components, as it is a dynamic quality that appears in varying degrees across speech and musical performance.

It was argued that one can distinguish among speech and music stimuli that have similar linguistic prosody or emotional prosody but bear other prosodic differences. These additional prosodic differences do not necessarily depend on syntactic structure, as syntax does not fully govern the prosodic realisation of a given utterance. Utterances with identical linguistic prosody and expressing the same emotions can still differ on other prosodic levels. A soloist can be also particularly flexible in the use of prosodic manipulations, which can go beyond prescribed syntactic norms. Analogous features in the choices of actors, which can be viewed as the counterpart aspect of 'expressiveness' in speech, differ among them and deviate from absolute prescriptive rules, especially when it comes to more talented

performers. Judgements pertaining to aesthetic appreciation, therefore, pertain to a new type of prosodic perception.

It should be noted that although 'expressiveness' is here argued to relate to aesthetic appreciation of a linguistic or a musical stimulus, it is in some ways distinct from the 'pleasantness' or 'unpleasantness' evoked by pathological speech quality. Studies of perception of pathological speech have also attempted to examine the effect of pitch variation, speech rate, and voice quality variations that were part of the manipulations employed in this study as well. Rating of pleasantness in dysphonic speech perception in pathological speech studies (e.g., Wolfe and Martin, 1997, Eadie and Doyle, 2005) differs from aesthetic appreciation examined here. In the present study, listeners were required to judge the aesthetic result of utterances and melodies based on differences that did not stem from phonatory impairments. In these studies, ratings might have been affected by perception of normality of speech production along with other factors and not by perception of artistic versus deadpan performances.

This prosodic aspect of 'expressiveness' constitutes an original contribution to the parallel study of language and music. Apart from defining it from a theoretical angle, this study also pointed to a dissociation between perception of 'expressiveness' across the domains of language and music. The next section refers to how 'expressiveness' was assessed in a neuropsychological framework.

5.7.3 Neural correlates of 'expressiveness'?

It is here argued that we are probably far from determining the neural correlates of 'expressiveness'. Investigations of other suprasegmental features such as linguistic

or emotional prosody that have taken place for a relatively long period of time have not systematically pointed to common and concrete neural substrates. Rather, conflicting evidence hitherto collected reveals significant grey areas concerning even issues of lateralisation. The studies mentioned below are, to some degree, relevant to 'expressiveness' and participants in them suffer from lesions similar to the patient tested in this study. These studies are treated only in an attempt to show the difficulty in determining possible neural correlates of perception of 'expressiveness'.

Impairments of prosody in language and music do not appear in exact patterns, especially if multiple types of prosodic performance are tested across different patients. The following examples suggest that prosodic impairments emerge in various forms. A focal left lesion in the superior temporal gyrus resulted in amusia and dysprosody in a non-professional musician in Piccirilli et al. (2000). Bautista and Ciampetti (2003) tested a patient with right temporo-occipital seizures resulting in a temporal monotonic speech output and also singing difficulty. Mendez (2001) describes that case of left handed patient who after suffering a right-temporo parietal stroke developed a particular appreciation for music becoming an avid concert attendee. At the same time, the patient displayed compromised speech and environmental sound perception and also had severe difficulties with melody naming. However, his prosody perception and production were intact. In a more thorough study of linguistic and musical prosody, Nicholson et al. (2003) report the case of a patient with right fronto-parietal damage that had a very poor performance on prosodic and music discrimination tasks, despite his ability to make successful judgements on segmental prosody tasks, suggesting possible overlapping networks for speech and music prosody.

One serious hurdle in determining these neural correlates is that the pathological consequences of a given neurological disorder varies greatly in the size and extent of damaged brain regions involved and other long-distance effects. This is the case with amusia, for example, which can be thought as a set of musical symptoms that, however, appear consequent to lesions in different brain areas. This hurdle can be even more significant if one thinks of the added complexity that every study can entail by designing new tasks not employed in previous studies or by using variations of previously used tasks. Differences in perception of emotional prosody, for example, might relate to differences in the task demands or to small differences in lesion profiles. Therefore, some systematic use of tasks in patients with similar neurological problems would possibly render this type of investigation more feasible in the future.

In more crude terms, looking at how 'expressiveness' is achieved across the domains of language and music is necessary in order to support the question of whether processing 'expressiveness' across domains might depend on the same networks. It seems that, although manipulation of the same acoustic features is required in order to produce 'expressive' speech and music output, some of them might be more important for language than music and vice versa. The tasks described in chapter 3 aimed to provide answers to these questions. The concept of 'expressiveness' can be empirically strengthened and validated by further research which identifies additional evidence of dissociation patterns. This can occur if future studies identify brain-damaged individuals that are able to perceive this prosodic dimension, while being impaired at perceiving linguistic and emotional prosody. Alternatively, if studies in the future identify patients with intact linguistic and emotional prosody

perception but compromised perception of 'expressive' prosody in speech and music performances, the concept of 'expressiveness' will become more robust.

5.8 Significance and Contribution

The present study added a further dimension to the parallel study of the prosodic organisation of speech and music. The main theoretical contribution to the field pertains to the suggestion that prosody can express features beyond pragmatics, syntax, and emotion. Evaluation of research on prosody revealed gaps in the existing prosodic descriptions and gave rise to the concept of 'expressiveness'. The empirical data of the study, thus, reveal an additional aspect pertaining to the relationship of 'expressiveness' across the linguistic and the musical domain. Appreciation of 'expressiveness' in speech might be independent of appreciation of 'expressiveness' in music, as suggested by the performance of the brain-damaged individual that participated in this study. Moreover, data from control participants and BZ reveal a reverse dissociation. That is, in contrast to the patient, control participants scored at ceiling when it came to 'expressiveness' in speech but their performance was not as good on the task examining perception of 'expressiveness' in music. This was the case even for controls that were matched with IB for music experience.

This study also contributes –to some extent– to the debate of 'modularity' of cognitive processing. Although the data is not sufficient to support modular organisation of speech prosody or music prosody, it suggests that despite their seemingly perceptual similarities, prosodic features of the two domains have been shown to dissociate and, hence, may be inferred to be processed separately, given the performance of IB and the other participants.

In terms of the methodological choices in the comparative study of language and music, it has been here shown that creating absolute stimuli counterparts across domains might not be the ideal way to examine the relationship between the two faculties. What is more, equal manipulations across domains might result in misleading conclusions. An initial stage of this investigation revealed that applying the manipulation that was used in the speech pitch task to music stimuli resulted in much greater difficulty for controls. As it was noted, a 2 semitone manipulation in the middle of a speech stimulus was easily perceived by control participants, whereas the same manipulation in music was not identified by the same listeners. This observation leads to the conclusion that accuracy in manipulations across domains does not guarantee the same amount of perceptual difficulty for the normal cognitive system. Consequently, in order to study speech and music impairments, one should ensure that dissociations do not emerge as a result of the technical manipulation of stimuli. In other words, dissociation resulting from different processing pathways should be considered different from dissociated performance due to different levels of difficulty between a speech and a music task due to an identical manipulation.

Another contribution of the study can be attributed to its suggestions for the further development of tools evaluating Amusia. Apart from the emotional aspect of music that is missing from the existing Amusia Battery, the Battery does not include tasks assessing judgments of the expressive quality of music. The new component of 'expressiveness' might also be fruitfully investigated in such individuals. Despite the fact that BZ had been considered congenitally 'amusic' in another study after completing the Amusia Battery, her performance on the 'expressive music prosody task' for the purposes of this study, suggests that this type of musical ability was not

compromised.   Hence, a person's inability to perform pitch judgements does not entail their inability to perform judgements of 'expressiveness' based on musical performance differences. Or, a person's perception of musical performance differences does not guarantee their ability to do well on the Montreal Amusia Battery tests.

The study may also have some implications for speech therapy and rehabilitation. As the brain-damaged individual of this study failed to perform speech prosodic discrimination judgements but had an impressively good performance on music prosody, one could suggest that music prosody tasks can be used in order to compensate for an individual's impaired perception of speech prosodic differences. This would, however, demand focused investigation of individual features in the music signal in order to detect the elements that could be transferable to the perceptual system of the speech domain. The present investigation raises the issue of rehabilitation through the use of tools, employing both speech and music in the presence of brain damage. Since the use of Melodic intonation therapy for aphasic patients (see Sparks and Holland, 1976), there has not been enough focus on the rehabilitation benefits that music can have on language. However, music is also likely to have a beneficial effect on rehabilitation of prosodic impairments that have received less attention in contrast to other components of language that seem to have a more 'obvious' contribution to communication.

In summary, this work has made a fourfold contribution to the study of speech and music. Firstly, its theoretical arguments and empirical data revealed an additional prosodic aspect, beyond the scope of syntax and emotion. Secondly, based on other tasks that were designed for its purposes, it raised a methodological issue in the parallel study of language and music; whether accuracy in manipulating individual

variables across domains guarantees valid comparisons in the perception in language and music. Thirdly, it suggested that the Montreal Battery of the Evaluation of Amusias could be enriched to include tasks on the appreciation of music's prosodic structure. Finally, the study may point to new avenues of development in the rehabilitation of prosodic impairments through the use of musical prosody, given that a patient might have impaired perception of speech prosody but intact perception of musical prosody as in the case of IB.

6. Limitations and Future Research

This short chapter refers to the limitations that this study has, mainly in terms of how the tasks were designed and how the results obtained from controls, IB, and BZ can be interpreted. At the same time, references are made to possible future experiments that are likely to cast more light on the grey areas of the present endeavour.

One possible limitation of the Battery is the fact that not all tasks employed the same response mode which possibly puts a constraint on appreciating the results of the tasks in comparison to each other. For instance, if the speech detection task included a same-different discrimination judgement response mode, a direct comparison to the expressive speech prosody task would be possible. This would facilitate examining to what extent the absence of lexico-semantic information affected the performance of participants across the two tasks. With respect to the speech rate task and the music tempo task, the existence of three possible answers makes these two tasks harder to compare with other tasks in terms of their degree of difficulty as their response mode was more demanding compared to the same-different discrimination judgements that were required on most other tasks.

Another limitation of the LMP Battery is that it does not provide an assessment of suprasegmental features perception in isolation in order to prove what the dissertation argues on a theoretical level. That is, it has been suggested that 'expressiveness' is a gestalt-like quality depending on the dynamic combination of several acoustic features. However, the Battery has not tested whether some acoustic features included in 'expressiveness' play a more important role compared to other cues. Every stimulus across domains included, at the same time, a number of acoustic cues, without accounting for the effect of every feature in isolation. Future experiments will be necessary in order to provide a data-driven answer to this question. If 'expressiveness' perception is impaired in the presence of spared pitch variation perception, loudness variation perception, length variation perception, pausing and grouping perception, the argument in favour of a dynamic prosodic percept combing the above features will be significantly strengthened.

The study showed that control participants of both the musically educated and the musically uneducated group had a similar performance across tasks and, therefore, a single control group was created. According to the results on expressive music prosody, the music training that the participants of the music group had received did not seem to have an effect on their ability to make expressive and non-expressive discrimination judgements about music performance. This finding gave rise to the argument that music training does not affect perception of appreciation of aesthetic features in music performance. The limitation of this argument lies in the fact that the present experiments did not control for different levels of music proficiency and different levels of 'expressiveness'. For example, individuals with 10 years of music training could have had a statistically significant difference in their performance.

Moreover, more subtle realisations of 'expressiveness' –which were not tested here– might indeed rely on music experience.

With respect to IB, it should be noted that any attempt to generalise results from his performance to all right hemisphere patients or more specifically right temporo-parietal stroke patients would not be legitimate. Therefore, one more limitation of this study is that its results reflect IB's 'idiosyncratic' performance. Plastic changes in the brain either as a result of rehabilitation or pre-stroke experience and possible training can have very intricate effects on performance. As far as IB's experience as an amateur musician is concerned, it might be hard to evaluate whether this resulted in an enhanced performance. Two possible explanations can be suggested. The first is that the stroke he suffered did not affect at all areas that might be responsible for processing features of 'expressiveness' in music. The second is that his music experience led to long-term plastic changes that allowed his perception of 'expressiveness' in music. Both scenarios would be reasonable but a definitive answer is not possible. In the neuropsychological context, one cannot, of course, test the same individual 'cancelling' the effect of music experience or music therapy. In contrast to artificial intelligence, for example, neuropsychology cannot grasp and test isolated variables as these described above but, at the same time, it constitutes an approach that does not try to simplify the intricacies of human cognition and rather delves into its genuine manifestations. This is, of course, a limitation inherent to the field in general and not specifically to this study.

It is not clear whether IB's music experience had an effect on his performance, as his musical profile as a drum player may not have been perfectly matched to controls. The controls with musical experience included those who had played a variety of instruments. However, it is still notable that there was no difference between those

controls with or without three years of musical experience on these tasks. The tasks were found to be relatively easy to perform regardless of musical training. The performance of IB on the speech pitch task raises some additional questions that were not addressed in the present investigation. The study did not include a task testing whether IB's perception of different voices or his recognition of familiar voices was intact. It remains unknown whether these abilities were preserved in IB. Additional testing would explore his ability to do perceptual discrimination tasks with voices. Voice recognition and unfamiliar voice discrimination are shown to relate to distinct types of processing (Van Lancker et al., 1989). Additional testing with IB could determine if processes of recognition and discrimination of voices would bear any relevance to the speech pitch task that was used in the study. That is, this additional piece of research would possibly shed some light on whether phonagnosia correlates to compromised perception of technical manipulations of a single voice. Data from this type of future investigation would probably fine-tune research on phonagnosia attempting to determine impairments of perceiving specific acoustic characteristics that account for phonagnosia. If dissociation between perceiving different voices and perceiving differences in the manipulated voice of a single speaker was observed, this would raise interesting questions for voice perception processes.

A limitation that should be acknowledged regarding the interpretation of BZ's performance on the tasks in the LMP Battery is that the neurotypical controls were not matched to her subject variables. The fact that BZ had a high level of education might account for her ability to successfully cope with the experimental demands generally and to concentrate on the questions of the tasks. The relatively lower educational profile of controls (matched to IB) might have reflected less experience

engaging with such artificial and cognitively demanding analytical tasks. However, it should be noted that the controls used in Paraskevopoulos et al. (2010) were not matched to her either. It would be desirable to test BZ with neurotypical controls matched for age, education, socioeconomic, and multilingual status. This could shed more light on her particular pattern of performance on the LMP Battery.

In summary, the answer format of the tasks and the lack of additional tasks to test possible further variables of 'expressiveness' constitute the limitations of the Battery that was designed for the study. With respect the performance of neurotypical participants, all conclusions related to the absence or presence of musical experience should be evaluated bearing in mind that results from these two groups do not necessarily apply to all musicians and non-musicians groups. Interpretation of IB's performance is also subject to some constraints that would possibly not exist in case of further testing. At the same time, as all single case studies, his case does not allow for any generalisation of performance patterns. Finally, BZ's picture would have also been clearer if she was compared to matched controls.

7. Conclusions

Within the general scope of the exploration of human auditory processing, this investigation proposed a novel aspect of comparison between speech and music prosody. The experiments that were conducted for the purposes of this study provided evidence supporting the argument that a new speech prosody dimension should be recognised. This new prosodic dimension was shown to dissociate from its counterpart percept in the music domain. At the same time, identical pitch manipulations across the domains of speech and music were shown to have different

effects on perception. The study also raised the issue of further evaluation of music perceptual processing, revealing a music ability that remained intact in the case of a brain-damaged individual with severe 'analogous' speech deficits and an individual with congenital amusia.

The prosodic label of 'expressiveness' was proposed in order to bridge the gap of linguistic, emotional, and pragmatic prosody categories, as it was argued that these do not constitute exhaustive descriptions of speech prosody. At the same time, 'expressiveness' was tested in two different manifestations; 'expressiveness' in speech and 'expressiveness' in music. As it was shown, at the presence of extensive right hemisphere damage, perception of expressive performance in the speech domain was compromised but it was preserved in the musical domain. This finding can have some implications pertinent to the concept of modularity. One could suggest that the percept of 'expressiveness' is modular, as mechanisms subserving its perception can dissociate in different cognitive domains. Hence, 'expressiveness' has been argued to bear similar characteristics in language and music but the neuropsychological dissociation reported here suggests that it can be broken down into perception of speech 'expressiveness' and perception of musical 'expressiveness'.

'Expressiveness' has been, therefore, argued to be modular across domains, but the questions remains as to whether prosodic streams are perceived as a whole or as multiple features processed separately. In this study, 'expressiveness' has been defined as a gestalt-like quality encompassing a dynamic combination of acoustic features, as it was hypothesised that it constitutes a single quality rather than a combination of isolable qualities. One can assume that there is an interdependence of acoustic features which imposes a 'combined' type of processing. That is, a given

acoustic cue will be processed differently in the absence or presence of other features in the same acoustic stimulus. For instance, isolating the effect of differences in loudness, duration, pausing, and articulation and studying them separately might not necessarily point to the same perceptual ability that is tested when one looks at these features combined in a single entity. In a hypothetical and non-data driven example, perception of duration in combination with loudness manipulations might be perceived differently from perception of duration in a context of manipulated articulation. The question of whether it is more legitimate to study acoustic features in isolation or in combination also pertains to the question of whether one wishes to study cognition using more naturally occurring versus more lab-oriented material. A combination of both is, of course, meaningful, but caution should be used when interpreting the applicability of the findings that arise in both research contexts.

In contrast to studies focusing on the cognitive benefit of music training on both music ability and cognition in general, this study suggested that there might be unexplored musical abilities that do not depend on the music training one might receive but rather on simple exposure to music throughout life. The specific example this study provides stems from the assessment of the perceptual ability of music prosody in healthy musically educated individuals, healthy musically uneducated participants as well as a brain-damaged patient with music experience and a congenital amusic individual. Conclusions regarding this 'universal' musical ability are derived from the performance of the two groups of healthy controls and the other two participants. As control participants belonging to either group could perform well on this task with no significant difference between groups, music training does not seem to have improved this ability. In other words, those who have not received music training can still distinguish aesthetically interesting performances from less

interesting ones, according to the data collected for this study. The performance of the right temporo-parietal brain-damaged patient does not offer direct evidence for the above point as he had music experience. However, it is still interesting that his extensive right-hemisphere lesion did not compromise this ability. What is more interesting, however, is that BZ, the congenital amusic individual that participated, was also able to perform successfully on the task. This raises the question of whether labelling someone as amusic is legitimate. The privative 'a' (the prefix indicating absence) in the word amusic suggests that if one is called amusic, they will not have any musical ability. Studies on amusia have shown that a small percentage of the population might suffer from congenital abnormal music perception. However, the term should be revisited, as it suggests an absolute lack of musical ability. The case of BZ, the congenital amusic that was still able to perform well on the expressive music prosody task of the present study, offers support for this argument. Appreciation of the aesthetic value of music does not seem to be compatible with the profile of 'amusia', that is, complete lack of musicality.

The study showed that methodological choices can have a direct effect on the assessment of perceptual abilities in the parallel investigation of language and music. The results from the exploration of pitch perception across domains in this study suggests that human auditory perception might be domain-specific, as a manipulation in one domain will cause perceptual differences of a different quality compared to another domain. In this specific case –from which this argument is drawn– a pitch manipulation in speech produced easily perceivable voice quality differences in contrast to the significant difficulty the same manipulation imposed when applied to music stimuli.

This study made an original contribution to the investigation of language and music, proposing an additional prosodic element across two domains that has not been studied before. At the same time, the design of other novel tasks raised methodological points regarding the way tasks across domains should be designed. This investigation contributes to the comparative study of language and music but also has specific implications for the study of music cognition as it questions some of the assumptions that batteries examining music abilities have made. Future research can respond to this and other similar questions, shedding more light on the intricacies of the language-music relationship and how this is instantiated in the human brain.

8. List of References

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology*, *29*, 130-131.

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E. & Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain*, 136, 1639-1661.

Alexander, J., Wong, P. C. M., & Bradlow, A. (2005). Lexical tone perception in musicians and nonmusicians. Paper presented at the Interspeech 2005 (Eurospeech) 9th European Conference on Speech Communication and Technology, Lisbon, September 2005.

Altenmuller, E. (2006). Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage*, 30, 917–926.

Anderson, W. (1973). Word-accent and Melody in Ancient Greek Musical Texts. *Journal of Music Theory*, 186-202.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, 16, 243-260.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *The American journal of psychology*, 147-166.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, *47*, 31-56.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, 125, 238-25.

Baird, A. D., Walker, D. G., Biggs, V., & Robinson, G. A. (2014). Selective preservation of the beat in apperceptive music agnosia: A case study. *Cortex*, 53, 27-33.

Ball, M. J., & Rahilly, J. (1999). *Phonetics: The science of speech*. Arnold.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., ... & Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage*, *30*, 917-926.

Baum, S. R., & Dwivedi, V. D. (2003). Sensitivity to prosodic structure in left-and right-hemisphere-damaged individuals. *Brain and Language*, *87*, 278-289.

Bautista, R. E. D., & Ciampetti, M. Z. (2003). Expressive aprosody and amusia as a manifestation of right hemisphere seizures. *Epilepsia*, 44, 466–467.

Besson, M., & Schön, D. (2003). Comparison between language and music, 269-294. In Peretz & R. Zatorre (eds.), *The Cognitive Neuroscience of music*. Oxford: Oxford University Press.

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative neurology and neuroscience*, *25*, 399-410.

Bidelman, G. M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *The Journal of Neuroscience*, *29*, 13165-13171.

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain research*, *1355*, 112-125.

Blonder, L. X., Bowers, D., & Heilman, K. M. (1991). The role of the right hemisphere in emotional communication. *Brain*, *114*, 1115-1127.

Blumstein, S. E., & Stevens, K. N. (1981). Phonetic features and acoustic invariance in speech. *Cognition*, *10*, 25-32.

Boersma, P. & Weenink, D. (1996). PRAAT: A system for doing phonetics by computer. Report of the Institute of Phonetic Sciences of the University of Amsterdam 132.

Bolton, T. (1894). Rhythm. *American Journal of Psychology*, 6, 145–238.

Boltz, M. & Jones, M. R. (1986). Does rule recursion make melodies easier to reproduce? If not, what does?. *Cognitive psychology*, *18*, 389-431.

Borod, J. C., Koff, E., Lorch, M. P., & Nicholas, M. (1986). The expression and perception of facial emotion in brain-damaged patients. *Neuropsychologia*, 24, 169-180.

Borod, J. C. (1992). Interhemispheric and intrahemispheric control of emotion: A focus on unilateral brain damage. *Journal of Consulting and Clinical Psychology*, 60, 339-348.

Botinis, A., Granstrom, B., & Mobius, B. (2001). Developments and paradigms in intonation research. *Speech Communication*, 33, 263–296.

Bowers, D., Coslett, H.B., Bauer, RM., Speedie, L., & Heilman, K.M. (1987). Comprehension of emotional prosody following unilateral brain damage: Processing deficit versus distraction deficit. *Neuropsychologia*, 25, 317-328.

Bowers, D., Bauer, R. M., & Heilman, K. M. (1993). The nonverbal affect lexicon: Theoretical perspectives from neuropsychological studies of affect perception. *Neuropsychology*, *7*, 433.

Brown, E., & Miron, M. S. (1971). Lexical and syntactic predictors of the distribution of pause time in reading. *Journal of Verbal Learning and Verbal Behavior*, 10, 658-667.

Budinger, E. (2005). Introduction: auditory cortical fields and their functions, 3-6. In Konig, Heil, Budinger & Scheich (eds.), *Auditory Cortex: A Synthesis of Human and Animal Research.* New Jersey: Lawrence Erlbaum Associates.

Callaway, C. B., & Lester, J. C. (2002). Narrative prose generation. *Artificial Intelligence*, 139, 213-252.

Caramazza, A. (1984). The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and language*, *21*, 9-20.

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain and cognition*, *5*, 41-66.

Caramazza, A., & Coltheart, M. (2006). Cognitive neuropsychology twenty years on. *Cognitive Neuropsychology*, *23*, 3-12.

Chater, N. (2003). How much can we learn from double dissociations? *Cortex*, *39*, 167-169.

Chen, A. (2005). *Universal and language-specific perception of paralinguistic intonational meaning*. Utrecht: LOT.

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in cognitive sciences*, *7*, 300-307.

Clark, J. E., Yallop, C., & 史宝辉. (2000). *An introduction to phonetics and phonology*. Oxford: Blackwell.

Coltheart, M. (2001). Assumptions and methods in cognitive neuropsychology. *The handbook of cognitive neuropsychology*, 3-21.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science*, 31, 697-704.

Crawford, J. R., & Garthwaite, P. H. (2007). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive Neuropsychology*, 24, 343-372.

Crawford, J.R. & Howell, D.C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, 12, 482-486.

Crystal, D. (2008). A dictionary of linguistics and phonetics. Oxford: Blackwell.

Cutler, A. (1974). On saying what you mean without meaning what you say. *Chicago Linguistic Society*, 10, 117-27.

Cutler, A. (1997). The comparative perspective on spoken-language processing. *Speech communication*, 21, 3-15.

Darby, D. G. (1993). Sensory aprosodia A clinical clue to lesions of the inferior division of the right middle cerebral artery?. *Neurology*, 43, 567-567.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant behavior and Development*, *9*, 133-150.

DeCheveigné , A. (2005). Pitch perception models, 169 – 233. In C. J. Place , A. J. Oxenham , R. R. Fay , and A. N. Popper (eds.). *Pitch—Neural Coding and Perception*. New York : Springer.

Doukhan, D., Rilliard, A., Rosset, S., Adda-Decker, M., & d'Alessandro, C. (2011). Prosodic Analysis of a Corpus of Tales. *Interspeech*, 3129-3132.

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H. G., Winkler, J., Büchel, C., & May, A. (2006). Temporal and spatial dynamics of brain structure changes during extensive learning. *The Journal of Neuroscience*, *26*, 6314-6317.

Eadie, T. L., & Doyle, P. C. (2005). Classification of dysphonic voice: acoustic and auditory-perceptual measures. *Journal of Voice*, 19, 1-14.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*, 210-233.

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition*, 100, 173–215.

Fitzsimons, M., Sheahan, N., & Staunton, H. (2001). Gender and the integration of acoustic dimensions of prosody: implications for clinical studies. *Brain and language*, *78*, 94-108.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain*, *127*, 801-810.

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain and language*, 89, 267-276.

Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (2007). Role of the corpus callosum in speech comprehension: interfacing syntax and prosody. *Neuron*, 53, 135-145.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X. & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody.*Neuroimage*, *23*, 344-357.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, 23, 9240-9245.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive psychology*, *15*, 411-458.

Gorelick, P. B., & Ross, E. D. (1987). The aprosodias: further functional-anatomical evidence for the organisation of affective language in the right hemisphere. *Journal of Neurology, Neurosurgery & Psychiatry*, 50, 553-560.

Gosselin, N., Jolicœur, P., & Peretz, I. (2009). Impaired memory for pitch in congenital amusia. *Annals of the New York Academy of Sciences*, 1169, 270-272.

Griffiths, T. D., Flees, A., & Green, G. G. (1999). Disorders of human complex sound processing. *Neurocase*, 5, 365-378.

Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive psychology*, 11, 58-81.

Gutschalk, A., Uppenkamp, S., Riedel, B., Bartsch, A., Brandt, T., & Vogt-Schaden, M. (2015). Pure word deafness with auditory object agnosia after bilateral lesion of the superior temporal sulcus. *Cortex*, 73, 24-35.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? S*cience*, *29*, 1569-1579.

Heilman, K. M., Scholes, R., & Watson, R. T. (1975). Auditory affective agnosia. *Journal of Neurology, Neurosurgery and Psychiatry*, 38, 69-72.

Heilman, K. M., Bowers, D., Speedie, L., & Coslett, H. B. (1984). The comprehension of emotional and non-emotional prosody. *Neurology* , 34, 917-921.

Herdener, M., Esposito, F., di Salle, F., Boller, C., Hilti, C. C., Habermeyer, B., & Cattapan- Hermes, D. J., & Rump, H. H. (1994). Perception of prominence in speech intonation induced by rising and falling pitch movements. *The Journal of the Acoustical Society of America*, *96*, 83-92.

Hutchins, S., Zarate, J. M., Zatorre, R. J., & Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *The Journal of the Acoustical Society of America*, *127*, 504-512.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, 15, 356-360.

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, *46*, 632-639.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music perception*, 26, 195-204.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2011). Fine-grained pitch discrimination in congenital amusics with Mandarin Chinese, *Music Perception: An Interdisciplinary Journal*, 519-526.

Johnsrude, I.S., Penhune, V.B. & Zatorre, R.J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain,* 123, 155–163.

Johnson, J. K., Graziano, A. B. & Hayward, J. (2010). Historical perspectives on the study of music in neurology, 17-30. In F. Clifford Rose (ed.), *The neurology of music*. London: Imperial College Press.

Johnson, J. K., Lorch, M., Nicolas, S., & Graziano, A. (2013). Jean-Martin Charcot's role in the 19th century study of music aphasia. *Brain*, awt055.

Jones, M. R. & Ralston, J. T. (1991). Some influences of accent structure on melody recognition. *Memory & Cognition*, 19, 8–20.

Jusczyk, P.W. & Krumhansl, C.L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 627–640.

Kaas, J. H., & Hackett, T. A. (2005). Subdivisions and connections of auditory cortex in primates: a working model, 7-25. In Konig, Heil, Budinger & Scheich (eds.), *Auditory Cortex: A Synthesis of Human and Animal Research.* New Jersey: Lawrence Erlbaum Associates.

Karmiloff-Smith, A. (1995). *Beyond modularity: A developmental perspective on cognitive science*. MIT press.

Kisilevsky, B. S., Hains, S. M. J., Jacquet, A. Y., Granier-Deferre, C., & Lecanuet, J. P. (2004). Maturation of fetal responses to music. *Developmental Science*, 7, 550-559.

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*, 13, 646-653.

Klouda, G. V., Robin, D. A., Graff-Radford, N. R., & Cooper, W. E. (1988). The role of callosal connections in speech prosody. *Brain and Language*, 35, 154-171.

Kolinsky, R., Cuvelier, H., Goetry, V., Peretz, I. & Morais, J. (2009). Music Training Facilitates Lexical Stress Processing. Music Perception, 26, 235-246.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y. & Friederici, A. D. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain and language*, *86*, 366-376.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11, 599-605.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, *97*, 11850-11857.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and cognition*, 69, 127-137.

Lancker, D. R. V., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *Journal of clinical and experimental neuropsychology*, 11, 665-674.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Massachusetts: MIT Press.

Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception*, 1, 229-252.

Liegeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V. & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain* 121, 1853–1867.

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain*, 133, 1682-1693.

Liu, F., Xu, Y., Patel, A. D., Francart, T., & Jiang, C. (2012). Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers. *Brain and cognition*, 79, 209-215.

Lorch, M. (2011). Re-examining Paul Broca's initial presentation of M. Leborgne: Understanding the impetus for brain and language research. *Cortex*, *47*, 1228-1235.

Lorch, M. P. (2013). Examining language functions: a reassessment of Bastian's contribution to aphasia assessment. *Brain*, *136*, 2629-2637.

Lorch, M. & Greenblatt, S. (2015). Singing in Speechless Children, 53-72. In E. Altenmuller, F. Boller and S. Finger (eds.) *Music, Neurology, and Neuroscience: Historical Connections and Perspective.*. Progress in Brain Research. Elsevier, New York.

Loutrari, A. and Lorch, M. (in press). 'Music and language expressiveness: when emotional character does not suffice'. To appear in: Penelope Gouk, Jacomien Prins, Wiebke Thormaehlen, and James Kennaway (eds.) *The Power of Music: Historical and Scientific Perspectives on Music, Emotions and Wellbeing*. Chicago University Press.

Ludewig, K. (2010). Musical training induces functional plasticity in human hippocampus. *The Journal of Neuroscience*, *30*, 1377-1384.

Luo, H., Ni, J. T., Li, Z. H., Li, X. O., Zhang, D. R., Zeng, F. G., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences*, 103, 19558-19563.

Marmel, F., Tillmann, B., & Dowling, W. J. (2008). Tonal expectations influence pitch perception. *Perception & Psychophysics*, 70, 841-852.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, 19, 1453-1463.

Mazzoni, M., Moretti, P., Pardossi, L., Vista, M., Muratorio, A., & Puglioli, M. (1993). A case of music imperception. *Journal of neurology, neurosurgery, and psychiatry*, 56, 322.

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current opinion in neurobiology*, 18, 452-463.

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception: An Interdisciplinary Journal*, 21, 289-31.

Mendez, M. F. (2001). Generalized auditory agnosia with spared music recognition in a left-hander. Analysis of a case with a right temporal stroke.*Cortex,* 37, 139-150.

Meyer, L.B (1956). *Emotion and meaning in music*. University of Chicago Press.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G. & von Cramon, D.Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, 17, 73– 88.

Meyer, M., Steinhauer, K., Alter, K., Friederici & von Cramon, D. Y. (2004) Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language,* 89, 277–289.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, 24, 167-202.

Nicholson, K.G., Baum, S., Kilgour, A., Koh, C.K., Munhall, K.G. & Cuddy, L.L (2003). Impaired processing of prosodic and musical patterns after right hemisphere damage. *Brain and Cognition*, 52, 382 – 389.

Nishimura, H., Hashikawa, K., Doi, K., Iwaki, T., Watanabe, Y., Kusuoka, H.,& Kubo, T. (1999). Sign language 'heard'in the auditory cortex. *Nature*, 397, 116-116.

Nygaard, L. C. (2008). Perceptual Integration of Linguistic and Nonlinguistic Properties of Speech, 390-414. In D. B. Pisoni & R. E. Remez (eds.) *The handbook of speech perception*, John Wiley & Sons.

Ogden, R. (2009). *An introduction to English phonetics*. Edinburgh University Press.

Oxford English Dictionary, 2015. "accent, n.". OED Online. September 2015. Oxford, Oxford University Press. http://www.oed.com.ezproxy.lib.bbk.ac.uk/view/Entry/989?rskey=zeBhKh&result=1 (accessed October 03, 2015).

Palmer, C. & Hutchins, S. (2006). What is musical prosody?. *Psychology of learning and motivation*, *46*, 245-278.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E. & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392, 811-814.

Paraskevopoulos, E., Tsapkini, K. & Peretz, I. (2010). Cultural aspects of music perception: Validation of a Greek version of the Montreal Battery of Evaluation of Amusias. *Journal of the International Neuropsychological Society*, *16*, 695-704.

Patel, A. D. (2003). Rhythm in language and music. *Annals of the New York Academy of Sciences*, *999*, 140-143.

Patel, A. D. (2008). Music, language, and the brain. New York: Oxford University Press.

Patel, A. D. (2012). Language, music, and the brain: a resource-sharing framework. *Language and music as cognitive systems*, 204-223.

Patel, A. D., Peretz I., Tramo, M. & Labreque R. (1998). Processing Prosodic and Musical Patterns: A Neuropsychological Investigation. Brain and Language, 61, 123-144.

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and cognition*, *59*, 310-313.

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, 119, 3034-3047.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception: An Interdisciplinary Journal,* 25, 357-368.

Pell, M. & Baum, S. (1997). The ability to perceive and comprehend intonation in linguistic and affective contexts by brain damaged adults. *Brain and Language*, 57, 80-99.

Pell, M. D. (2006). Cerebral mechanisms for understanding emotional prosody in speech. *Brain and language*, 96, 221-234.

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, *6*, 661-672.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, 113, 1185-1205.

Peretz, I., Kolinsky, R. G., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, 117, 1283-1301.

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, 100, 1–32.

Peretz, I., Ayotte, J., Zatorre, R.J., Mehler, J., Ahad, P., Penhune, V.B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, 33, 185-191.

Peretz, I. & Colheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6, 688-691.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in cognitive sciences*, 7, 362-367.

Peretz , I. , Champod , A.S. & Hyde , K. (2003). Varieties of musical disorders, the Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Science* , *999* , 58 – 75.

Piccirilli, M., Sciarma, T., & Luzzi, S. (2000). Modularity of music: evidence from a case of pure amusia. *Journal of Neurology, Neurosurgery & Psychiatry*, 69, 541-545.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech communication*, *41*, 245-255.

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A. S., McNamara, J. O. & Williams, S. M. (2004). *Neuroscience* (3rd edition).*Sunderland, MA: Sinauer Associates*.

Roach, P. (1982). On the distinction between 'stress-timed'and 'syllable-timed'languages. *Linguistic controversies*, 73-79.

Ross, E. D. (1981). The aprosodias: Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology*, *38*(9), 561-569.

Ross, E. D., Thompson, R. D., & Yenkosky, J. (1997). Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain and language*, 56, 27-54.

Ross, E. D., Orbelo, D. M., Cartwright, J., Hansel, S., Burgard, M., Testa, J. A., & Buck, R. (2001). Affective-prosodic deficits in schizophrenia: profiles of patients with brain damage and comparison with relation to schizophrenic symptoms. *Journal of Neurology, Neurosurgery & Psychiatry*, *70*, 597-604.

Ross, D., Choi, J. & Purves, D. (2007). Musical intervals in speech. *Proceedings of the National Academy of Sciences*, 104, 9852-9857.

Ross, E. D., & Monnot, M. (2008). Neurology of affective prosody and its functional–anatomic organization in right hemisphere. *Brain and language*,*104*, 51-74.

Repp, B. H. (1992). A constraint on the expressive timing of a melodic gesture: Evidence from performance and aesthetic judgment. *Music Perception*, 221-241.

Saffran, E. M. (1982). Neuropsychological approaches to the study of language. *British Journal of Psychology*, *73*, 317-337.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.

Sammler, D., Kotz, S. A., Eckstein, K., Ott, D. V., & Friederici, A. D. (2010). Prosody meets syntax: the role of the corpus callosum. *Brain*, 133, 2643-2655.

Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: tests of children and adults. *Journal of Experimental Psychology: General*, 131, 511.

Scherer, K. R. (1988). On the symbolic functions of vocal affect expression. *Journal of Language and Social Psychology*, 7, 79-100.

Scherer, K. R., Ladd, D. R., & Silverman, K. E. (1984). Vocal cues to speaker affect: Testing two models. *The Journal of the Acoustical Society of America*, 76, 1346-1356.

Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, 33, 1047-1055.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language.*Psychophysiology*, 41, 341-349.

Schönwiesner, M., von Cramon, D. Y., & Rübsamen, R. (2002). Is it tonotopy after all?. *Neuroimage*, 17, 1144-1161.

Schreuder, M. (2006). *Prosodic processes in language and music.* (Doctoral dissertation).

Schuppert, M., Münte, T. F., Wieringa, B. M., & Altenmüller, E. (2000). Receptive amusia: evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain*, 123(3), 546-559.

Schwartz, M. F. (1984). What the classical aphasia categories can't do for us, and why. *Brain and language*, 21, 3-8.

Schwartz, D. A., Howe, C. Q., & Purves, D. (2003). The statistical structure of human speech sounds predicts musical universals. *The Journal of Neuroscience*, 2318, 7160-7168.

Schwartz, S. R., Cohen, S. M., Dailey, S. H., Rosenfeld, R. M., Deutsch, E. S., Gillespie, M. B.,& Patel, M. (2009). Clinical practice guideline: hoarseness (dysphonia). *Otolaryngology-head and neck surgery*, 141, 1-31.

Schwartz, M. F., & Dell, G. S. (2010). Case series investigations in cognitive neuropsychology. *Cognitive neuropsychology*, 27, 477-494.

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, *23*, 450-466.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in neurosciences*, 26, 100-107.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press.

Shapiro, B. E., & Danly, M. (1985). The role of the right hemisphere in the control of speech prosody in propositional and affective contexts. *Brain and language*, 25, 19-36.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research*, 25, 193-247.

Skandera, P., & Burleigh, P. (2005). A manual of English phonetics and phonology. *Tübingen: Gunter.*

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *The quarterly journal of experimental psychology*, *35*, 377-396.

Snow, D. & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, 112, 1025–1058.

Sparks, R. W., & Holland, A. L. (1976). Method: melodic intonation therapy for aphasia. *Journal of Speech and Hearing Disorders*, *41*, 287-297.

Stein, J.F. & Stoodley, C.J. (2006). *Neuroscience: An Introduction*. West Sussex: John Wiley and sons.

Steinke, W.R., Cuddy, L.L. & Jakobson, L.S. (2001). Dissociations among functional subsystems governing melody recognition after right-hemisphere damage. *Cognitive. Neuropsychology*, 18, 411–437.

Stewart, L., von Kriegstein, K., Warren, J.D. & Griffiths, T.D. (2006) Music and the brain: disorders of musical listening. *Brain*, 129, 2533–2553.

Theune, M., Meijs, K., Heylen, D. & Ordelman, R. (2006). Generating expressive speech for storytelling applications. *Audio, Speech, and Language Processing, IEEE Transactions on*, *14*, 1137-1144.

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and cognition*, 71, 259-264.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychological review*, *107*, 885.

Todd, N. (1985). A model of expressive timing in tonal music. *Music perception*, 33-57.

Vaissière, J. (2008). Perception of Intonation , 236-263. In D. B. Pisoni & R. E. Remez (eds.) *The handbook of speech perception*, John Wiley & Sons.

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech communication*, *48*(9), 1162-1181.

Vos, P. G., & Troost, J. M. (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception*, 383-396.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & psychophysics*, *64*, 198-207.

Warrier, C. M., & Zatorre, R. J. (2004). Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain*, *127*, 1616-1625.

Wertz, R. T., Henschel, C. R., Auther, L. L., Ashford, J. R., & Kirshner, H. S. (1998). Affective prosodic disturbance subsequent to right hemisphere stroke: A clinical application. *Journal of Neurolinguistics*, *11*(1), 89-102.

White-Schwoch, T., Carr, K. W., Anderson, S., Strait, D. L., & Kraus, N. (2013). Older adults benefit from music training early in life: biological evidence for long-term training-driven plasticity. *The Journal of Neuroscience*, 33, 17667-17674.

Wightman, C. W. & Ostendorf, M. (1994). Automatic labelling of prosodic patterns. *IEEE Transations on Speech and Audio Processing*, 2, 469–481.

Wichmann, A. (2002) Attitudinal intonation and the inferential process. *Speech Prosody 2002 International Conference*, Aix-enProvence.

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory*, 18, 657-669.

Wolfe, V., & Martin, D. (1997). Acoustic correlates of dysphonia: type and severity. *Journal of communication disorders*, *30*, 403-416.

Wong, P.C., Skoe E., Russo, N.M., Dees T. & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10, 420-422.

Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., ... & Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, 27, 173-183.

Yuan, J., Shih, C., & Kochanski, G. P. (2002). Comparison of declarative and interrogative intonation in Chinese. In *Speech Prosody 2002, International Conference*.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256, 846-849.

Zatorre R.J., Evans, A.C. & Meyer E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience*, 14, 1908–1919.

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences*, 95, 3172-3177.

Zatorre, R.J., Belin, P. & Penhune, V.B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6, 37-46.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8, 547-558.

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biol*, *10*, e1001372.

Zimmermann, M. (2007). Contrastive focus, 147-159. In Féry, C., G. Fanselow, and M. Krifka (eds.), *The Notion of Information Structure*. University of Potsdam, Humboldt University of Berlin, Germany.

APPENDIX

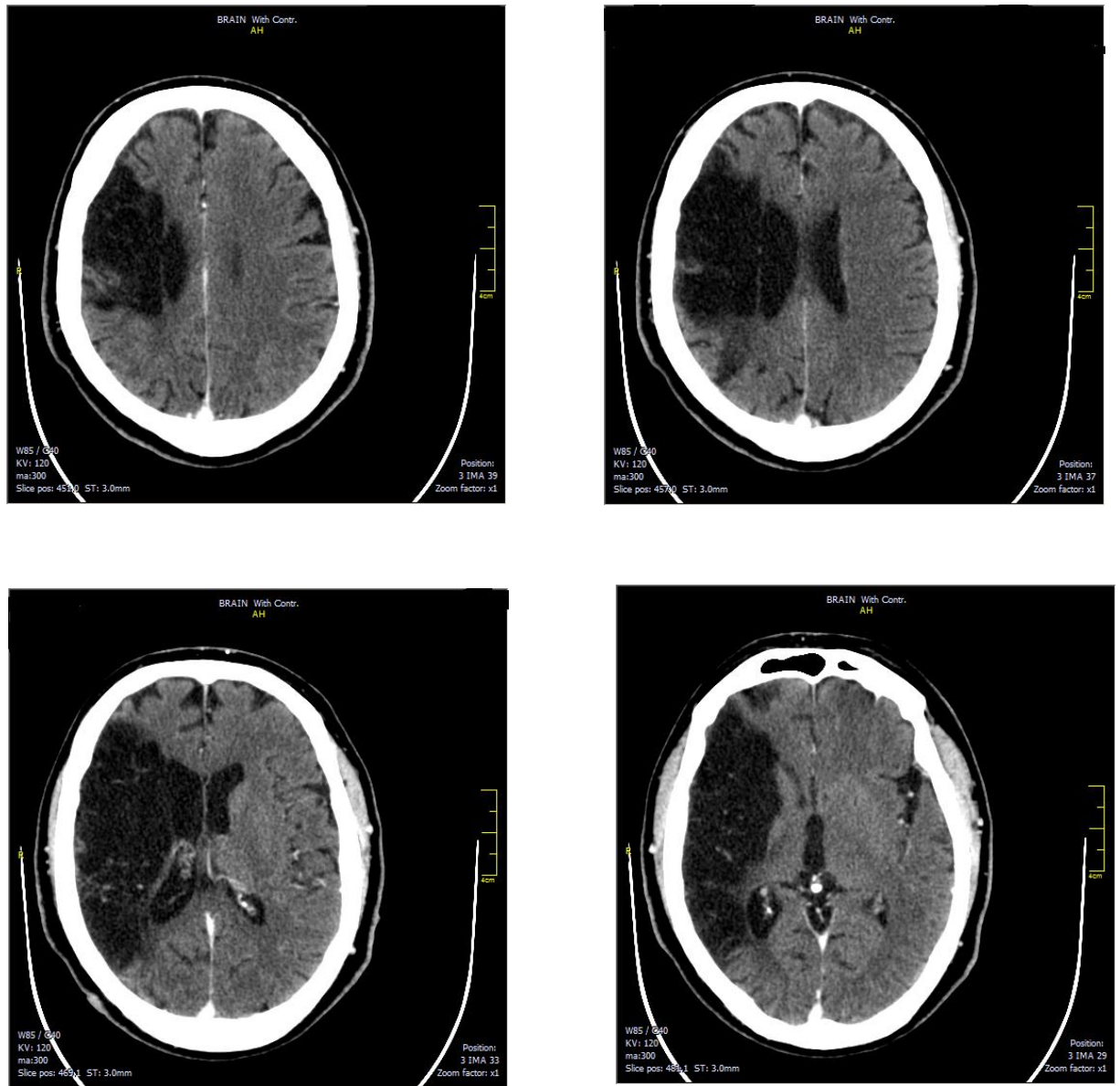Representative IB scans are shown in Figure 22.

Figure 22: Transverse CT scans obtained from IB 9 months post-stroke. The right side of the brain is shown on the left side of the scan. The scans show extended damage in the right temporo-parietal lobe. The scans proceed inferiorly, starting from the image appearing in the upper left-hand corner.

Controls	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
1	28/32	20/32	30/32	31/32	30/32	27/32	30/32	19/32
2	21/32	27/32	31/32	32/32	27/32	26/32	25/32	17/32
3	32/32	24/32	31/32	32/32	30/32	29/32	32/32	24/32
4	29/32	21/32	31/32	32/32	31/32	31/32	31/32	24/32
5	30/32	26/32	31/32	32/32	32/32	29/32	31/32	25/32
6	30/32	21/32	25/32	32/32	31/32	26/32	31/32	26/32
7	32/32	29/32	31/32	32/32	31/32	31/32	32/32	29/32
8	32/32	27/32	29/32	32/32	31/32	18/32	32/32	19/32
9	31/32	27/32	32/32	32/32	32/32	20/32	28/32	26/32
10	31/32	23/32	29/32	30/32	27/32	19/32	28/32	24/32
11	32/32	24/32	29/32	32/32	32/32	29/32	32/32	28/32
12	25/32	17/32	31/32	31/32	27/32	27/32	30/32	28/32
13	32/32	27/32	31/32	32/32	32/32	32/32	30/32	27/32
14	32/32	24/32	31/32	32/32	32/32	32/32	32/32	29/32
15	31/32	28/32	31/32	32/32	31/32	30/32	32/32	26/32
16	29/32	25/32	31/32	32/32	27/32	31/32	32/32	13/32
17	30/32	22/32	31/32	32/32	32/32	32/32	30/32	16/32
18	26/32	22/32	29/32	32/32	31/32	28/32	30/32	25/32
19	32/32	25/32	31/32	32/32	32/32	32/32	32/32	23/32
20	24/32	22/32	28/32	32/32	31/32	31/32	31/32	22/32
21	32/32	20/32	29/32	31/32	30/32	31/32	31/32	23/32
22	32/32	29/32	30/32	32/32	32/32	31/32	29/32	22/32
23	20/32	27/32	31/32	32/32	31/32	29/32	32/32	21/32
24	19/32	23/32	31/32	32/32	27/32	31/32	31/32	23/32

Table 5: Performance of controls on the 8 tasks of the LMP Battery: Task 1 (speech prosody detection), Task 2 (expressive music prosody), Task 3 (expressive speech prosody), Task 4 (emotional speech prosody), Task 5 (emotional music prosody), Task 6 (speech pitch), Task 7 (speech rate), Task 8 (music tempo).

Task Instructions

The following instructions were given to participants before the presentation of practice items. In case a participant was not clear on the instructions after the completion of practice items, instructions were repeated.

For the speech prosody detection task, participants were informed that they would be exposed to pairs of auditory stimuli for each question. They were instructed to attend carefully but to not compare the stimuli. It was explained to all participants that some of the sounds would possibly remind them of a person speaking behind closed doors and some other sounds would more similar to noise rather than human speech. Also, detailed instructions were provided with respect to the response mode. Listeners were informed that they could tick the option 'the first stimulus is a speech stimulus' or the option 'the second stimulus is a speech stimulus' or 'both stimuli are speech stimuli' or 'neither of the stimuli is a speech stimulus'. It was made clear to them that if they ticked the first two answers ('the first stimulus is a speech stimulus' and 'the second stimulus is a speech stimulus') their answer would equal to 'Both stimuli are speech stimuli' and that they would not have to untick the first two boxes and then tick the third one.

The instructions on the 'expressive music prosody task' were as follows: the test includes melodies played on a piano displaying a major difference; some of them are played by a musician and some of them are played by a machine. If the participants thought that both melodies were executed by a musician or by a machine, then the correct answer would be 'same'. By contrast, if one of them was played by a musician and the other was played by a machine, the correct answer would be 'different'. Listeners were instructed to be careful and only mark one out of two possible answers. A hint that was given to them as part of the instructions was that the melodies played by a musician would sound more expressive and those played by a machine would sound more artificial.

For the expressive speech prosody detection task, participants were informed that they would listen to recordings of a male actor saying his lines. They were instructed to attend to the way he said the lines and decide if he was equally expressive in both stimuli they were presented with. It was made clear to the participants that the words included in every pair of sentences remained exactly the same. However, if they found any difference in the way the actor uttered the lines, they should tick the answer 'different'. If they thought that both stimuli were exactly the same, they were instructed to choose the answer 'same'.

On the emotional speech prosody task, participants were instructed to listen to individual stimuli for each question and choose between two possible answers, namely 'happy' or 'sad'. They were informed that the words that were contained in the sentences did not relate to the emotion that they had to identify. Rather, they

were encouraged to ignore the words of the sentences and focus on the emotion in the voice of the speaker. It was explained to them that only one answer would be possible.

The instructions for the emotional music prosody task were similar. Participants were told that they were required to mark melodies as 'happy' or 'sad'. It was made clear that no lyrics would be present in the melodies and they would only need to base their judgement on whether they thought that they melodies created a happy or a sad mood. It was reminded to them that, as in the previous task, they would only listen to one stimulus at a time and that only one answer was possible.

The speech pitch task instructions were as follows: a pair of sentences will be presented for each question. There are two possible answers: either the sentences should be marked as 'same' or as 'different'. The listener should attend carefully to the second half of each sentence because it is the second half of a sentence that shows if the sentence is same or different from the previous sentence of a given pair. More specifically, the listener was instructed to focus on the voice of the speaker and decide whether the voice of the speaker was identical or different in the second half of the stimuli. It was made clear to the listener that there would be no difference in terms of the words that were included in the sentences.

For the speech rate task, participants were instructed to attend to the spoken stimuli and decide whether they were equally fast or not. If they thought that the speed of both stimuli in a pair was identical, they were required to choose the answer 'the utterances were equally fast'. If there was a difference in rate, they were instructed to indicate which utterance specifically was faster choosing between 'first utterance was faster' or 'second utterance was faster'.

Finally, on the music tempo task, listeners were instructed again to choose among three answers; 'first melody was faster', 'second melody was faster', 'the melodies were equally fast'. They were reminded that only one answer was possible and that they had to listen to both melodies before ticking the correct box.