Sounding sparks

Incongruities and aesthetics in music and language

Author:
Cara R. Featherstone

Supervisors:
Dr. Mitch G. Waterman
Dr. Catriona M. Morrison

The University of Leeds
Institute of Psychological Sciences

Submitted in accordance with the requirements for the degree of Doctor of Philosophy
September 2010

Intellectual property and publications statement

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored manuscripts has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that her contribution was intellectual and that of the co-authors, Dr. Mitch Waterman and Dr. Catriona Morrison, was primarily editorial.

Chapter 2 (the systematic review) led to the following manuscript to be submitted shortly for publication:

- Featherstone, C.R., Morrison, C.M. & Waterman, M.G. (in prep. for *Psychology of Aesthetics, Creativity and the Arts*) Reviewing aesthetic incongruities: a case for the role of closure in affective responses to music and language.

Elements from Chapters 3 (the description of the stimulus set), 4 (the outcome of the rating study), Chapter 5 (the stimulus norming procedure) and Chapter 7 (the oddity data from the ERP study) led to the following manuscript to be submitted shortly for publication:

- Featherstone, C.R., Waterman, M.G. & Morrison, C.M. (in prep. for *Behavior Research Methods*) Norming the Odd: Development, selection and validation of a stimulus set for the study of aesthetic incongruities across music and language.

The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Acknowledgements

This research, funded by an ESRC Open Competition Scholarship, would not have been possible without the contribution of many people along the way. First and foremost, I would like to thank my supervisors, Mitch Waterman and Catriona Morrison for their inspiration and support throughout this research. It's been a pleasure working with you and I am very grateful for the way in which you guided me through these three years. I'd also like to thank Lucy MacGregor for helping me get to grips with ERP methods, Karrie Alexander for applying her programming skills to my slightly mad ideas, Naomi Tomkins for being the voice of my experiments, Patrick Sturt, Daniela Sammler and other experts in the field for their kind advice in times of data-induced confusion. My thanks go also to the participants who took part in my experiments, and to the support staff of the Institute of Psychological Sciences, particularly Jackie, Julie and Kathryn, whose smiles, encouragement, and dedication contribute to the happy memories of my time in this department. On that note, I'd like to thank Nat, Tom, Amrit, Sarah, Ben, Erika, James, Vicky, in no particular order, for making these three years ones of great friendships as well as hard work.

I would also like thank my friends and family for their love and support. Some of you will recognise yourselves in my stimulus set. Karen, Sarah, Fawcett, Caroline, Rhian, Nan, Shannon, Andy, Sophie, Tomkins, Olivia, Kim, Jenny, Lucy, Lou, Gary, Jacquie, Pixi, Bill, Gill, David, Margaret, Hannah, Matthew, thank you for keeping me going through the tough times and for making the fun ones even better. I'd particularly like to thank my parents, for moving to a different country when I was young enough to learn to speak a new language, for always letting me play the various instruments that were lying around the house, even when I was too small to see the keys I was pressing on the piano, and for passing their love of music and language on to me. Thank you for supporting me throughout this endeavour in so many ways. A special thought goes also to my maternal grandparents, Joan and John Sherwood, whose quiet generosity saw me through my undergraduate studies, and whose unpretentious thirst for knowledge remains an inspiration for me. It's a shame you never got to see this. I like to think you would have said you found it “rather delicious”. Finally, a special thank you goes to Andrew Thompson, my Ein Gedi and my husband-to-be, whose love and support point me each day to the One to whom my greatest thanks will ever be due.

Abstract

Theories regarding the relationship between incongruities and aesthetic-affective responses show striking similarities across music and language. Traditional accounts as well as more recent theoretical and methodological advances suggest a model in which an element which is incongruous with expectations contributes to a heightened aesthetic experience. However, to date there has been no systematic comparison of the processing of aesthetically-purposed incongruities across these two domains of human cognition. Furthermore, within both music and language, researchers are calling for a more unified approach and a more systematic investigation of incongruity-based theories of aesthetics. The research carried out in this thesis aimed to investigate whether a shared model of incongruities and aesthetics could account for the relationship between incongruities and aesthetic-affective responses to music and language by collecting both behavioural and ERP data in experiments using a purpose-built stimulus set presenting equivalent levels of incongruity across the four components of music and language: harmony, rhythm, semantics and syntax. The results showed striking similarities between the patterns seen in behavioural and ERP data across *harmony* and *semantics*, supporting the notion that, in both components, incongruities do play a similar role in aesthetic-affective responses. This research also enabled the reconciliation of different patterns observed in seminal studies in the field of music and language ERP research, as well as further insights into the interpretation of ERP components through the comparison of behavioural data and ERP effects. Differences observed between the responses of musicians and non musicians to both music and language stimuli, the most striking of which was the absence of an oddity-related ERP component timelocked to semantic incongruities, support the notion that musical training affects other aspects of human cognition and suggest avenues for future research into the effects of musical training on language processing. Thus, alongside the main finding of a similar effect of aesthetically-purposed incongruities in music and language, the rigourous approach adopted in this investigation resulted in additional findings with equally important implications for the study of music and language psychology.

Contents

List of Figures

List of Tables

Chapter 1

General introduction: music, language, incongruities and aesthetics

Music: breathing of statues. Perhaps:

silence of paintings. You language where all language

ends. You time

standing vertically on the motion of mortal hearts.

- Rainer Maria Rilke (1875 - 1926)

1.1 Music, language, incongruities and aesthetics

1.1.1 Incongruity-based approaches to aesthetic-affective responses to music

Meyer's (1956) *Emotion and meaning in music* presented a model of musical communication which held the following central hypothesis:

> Affect or emotion-felt is aroused when an expectation — a tendency to respond — activated by the musical stimulus situation, is temporarily in-

hibited or permanently blocked (p. 31)

This model, which is considered the most robust model of emotions and meaning in music to date (Juslin & Västfjäll, 2008), is founded upon two theories concerning human perception. The first is a striving for clarity and stability. Meyer's (1956) model states that a situation which is structurally weak or organisationally unclear creates a tendency towards clarification. This conceptualisation of human perception echoes the Gestaltists' law of Prägnanz, which states "of several geometrical possible organisations that one will actually occur which possesses the best, simplest and most stable shape" (Koffka, 1935, p. 138). Subsequent studies in visual perception have repeatedly demonstrated this phenomenon (Pomerantz, 1981; Rock & Mack, 1994; Spillmann, 2006) as well as the top-down influence of prior knowledge on the perception of structure in a visual array (Geisler, Perry, Super, & Gallogly, 2001; Pomerantz, 1981; Rock & Palmer, 1990; Sekuler & Blake, 2002; Vecera & Farah, 1997).

Although much of the Gestalt literature focuses on the perception of two-dimensional visual objects, its principles have been applied to auditory perception, most notably in the area of auditory scene analysis (ASA) (Bregman, 1990). ASA is the process whereby a listener disentangles and integrates overlapping sounds from various sources over time, allowing for the recognition of footsteps and the comprehension of continuous speech (Sussman, 2005). The organisation of an auditory scene into distinct streams and the integration of sequential elements to perceptual units within each stream has been portrayed as underlying our ability to process two concurrent speech messages (Rivenez, Guillaume, & Darwin, 2006), to appreciate music (Sussman, 2005), and to hear interleaved melodies as separate entities (Dowling & Harwood, 1986).

Sekuler and Blake (2002) referred to the "occurrence statistics of images encountered in the natural world" (p. 211) in explaining the top-down influence of prior knowledge derived from past experience on the perception of visual contour. The top-down influence of implicit knowledge of musical structure, forms the second foundation upon which Meyer's (1956) hypothesis stands. Past experience here includes both the immediate past of the stimulus, which suggests probable continuations within the same stimulus, and past experiences of other similar stimuli, resulting in a set of expectations based upon probabilities within a particular style of music. A study by Tillmann, Bharucha, and Bigand (2000) demonstrated listeners' ability to evaluate novel stimuli according to rules which they implicitly learnt through mere exposure to music constructed according to finite grammar, lending support to Meyer's (1956) theorisation of the development of expectations. These sources of knowledge govern the way in which the listener organises musical elements into patterns, and lead to expectations as to the continuation of an unfolding pattern.

The addition of the organisational influence of prior knowledge to the striving for stability and clarity described by the Gestaltists underpins models of expectations in music processing. Through past experience, the listener acquires a mental representation of probabilities for a style of music previously encountered and for the likely organisation and progression of musical elements within a given piece. Meyer's (1956) model claims that upon hearing an unexpected element, the listener strives to integrate it into the unfolding piece, in accordance with the probabilities present both within the style of the piece and derived from the previous understanding of the piece. In a mechanism akin to that later described in the cognitive theory of emotions (Dowling & Harwood, 1986), expectations as to how incongruous elements will be integrated result in ease and unease, tension and relaxation, satisfaction and desire, and pleasure and pain, when they are thwarted through delays, deceptive diversions, unexpected consequents and ambiguous antecedents (Budd, 1985; Krumhansl, 2002; Meyer, 1957). Recent neuroimaging research has suggested several mechanisms underlying these emotions. Friston (2005) suggested a model in which "the comparison between prediction and actual input produces an error term which, if sufficiently large, will try to force an update of the model" (Vuust and Kringelbach, 2010, p. 173). The model developed by Huron (2006) posits that musical expectations, which are constantly changing as a listener progresses through a piece (Vuust & Kringelbach, 2010), lead to predictions as to what will follow. Predictions which prove correct, such as the prediction of a delayed tonic in a dominant-tonic cadence, are rewarded, leading to positive emotions. This approach to musical pleasure is supported by the involvement of reward circuitry in the experience of musical chills (Blood & Zatorre, 2001) and the involvement of paralimbic regions in the processing of dissonant sounds (Blood, Zatorre, Bermudez, & Evans, 1999).

Though the specific mechanisms remain uncertain and in need of further research (Juslin & Västfjäll, 2008; Vuust & Kringelbach, 2010), authors agree that the resolution of the tension caused by this striving, which confirms a previously destabilised prediction or brings the piece to a new stage of meaning, forms the basis of the affective response to music.

1.1.2 Echoes in theories of language aesthetics

Although the referential nature of language does not require it to resort to the mechanisms described above for the communication of meaning, the preface to Meyer's (1956) book suggests a "striking similarity of some aspects of musical experience to other types of aesthetic experience, particularly those evoked by literature" (p. ix). Echoes of his model of affective aesthetic musical experience can be seen in theories concerning the aesthetic and affective properties of figurative language.

Ortony (1975) claimed that metaphors were a necessary part of language, allowing the description of phenomenal experience with a vividness beyond what could be achieved by literal language. This vividness arises through the juxtaposition of incongruous elements. Considered by some as fundamental to human cognition (Noveck, Bianco, & Castry, 2001), metaphors have received a renewed interest over the past two decades (Bowdle & Gentner, 2005). Recent research into the processing of puns (Coulson & Van Petten, 2002), jokes (Coulson & Van Petten, 2002; Ritchie, 2005; Tartter, Gomes, Dubrovsky, Molholm, & Stewart, 2002) and the Shakespearean functional shift (e.g. "he *godded* me", Thierry et al., 2008) has extended the study of how language can be used to reflect a speaker's intentions above and beyond what is being expressed (Katz, 2005) to other forms of figurative language (Blasko & Kazmerski, 2006; Coulson & Van Petten, 2002, 2007; Tartter et al., 2002).

Theories of metaphor processing can be traced back to the Greek philosopher Aristotle (384-322 BC) who laid the foundations of the *Feature Mapping* theory by arguing that a metaphor was essentially an alternative grammatical construction of a simile. This theory suggests that understanding is achieved through the search for correspondences between the disparate domains of knowledge presented by the topic and the vehicle of the metaphor (Bowdle & Gentner, 2005). Although this theory has been the object of a series of revisions, its central tenet still holds, namely that the extra vividness of figurative language is generated via a striving for the integration of two distinct elements into a meaningful unit (Bowdle & Gentner, 2005). The role of the detection and integration of incongruities has been demonstrated in other forms of figurative language, forming the crux of many explanations of humour (Coulson & Williams, 2005; Norrick, 2003; Ritchie, 2005; Yus, 2003) and aesthetic emphasis (Clark & Clark, 1979; Thierry et al., 2008).

1.1.3 Methodological and theoretical developments in the study of affective aesthetic responses to music and language

One of Meyer's (1956) concerns with regards to the testing of his theory of musical experience was the difficulty in pinpointing the precise event in an unfolding musical sequence which led to the affective response described by listeners in introspective studies. The development over the last two decades of the Event-Related Potential (ERP) technique has given rise to a renewed interest in the study of incongruity processing (both their detection and integration). This technique, which uses Electroencephalography (EEG) recordings time-locked to specific events, enables the investigation of neural electricity patterns which reach the scalp of participants at precise moments in time. Its methodology relies upon the assumption that "if the average ERPs from two experimental conditions differ reliably at any given point in time, it can then be inferred that the associated brain and mental activity also differ at least by that point" (Kutas, 1998, p. 955).

The use of ERPs in the study of music and language processing has enabled psychologists to determine the electrical patterns recorded on the scalp of individuals upon occurrence of a specific chord (Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Koelsch, Grossmann, et al., 2003; Jentschke, Koelsch, & Friederici, 2005; Maess, Koelsch, Gunter, & Friederici, 2001), note (Besson & Macar, 1987; Hantz, Kreilick, Kananen, & Swartz, 1997; Miranda & Ullman, 2007), word (Coulson & Van Petten, 2002, 2007; Tartter et al., 2002; Thierry et al., 2008) and syllable (Van Berkum, Brown, Zwitserlood, Koojman, & Hagoort, 2005). Such studies have served to establish specific ERP patterns associated with violations of the rules of semantics and syntax (Osterhout, Allen, McLaughlin, & Inoue, 2002), harmony (Besson & Macar, 1987) and rhythm (Jongsma, Desain, & Honing, 2004). Thus, the identification of the precise element to which the affective response is tied seems more within reach nowadays than at the time of Meyer's (1956) initial publication. In a later publication, Meyer (1974) put forward the notion that "a solid empirical basis for a theory of the arts will not be established until the biochemical functioning of the nervous system can be related to human cognitive and affective behaviour" (p. 34). Despite the advances of ERP methods, allowing researchers to precisely determine the physiological pattern associated with a specific musical or linguistic element, the relationship between the elements, the physiological patterns and the affective experience are still unknown (Thierry et al., 2008; Juslin & Västfjäll, 2008). Furthermore, the interpretation of the observed ERP patterns themselves remains problematic.

In the language literature, ERP patterns in response to metaphor have revealed that the vehicle (B in a sentence of the form *this A is a B*) elicits an N400, a negative deflection occurring approximately 400ms after the onset of a target element. Evidence suggests an inversely proportional relationship between the amplitude of an N400 component and the degree of fit of a word in its semantic context (Lau, Phillips, & Peoppel, 2008; Nieuwland & Van Berkum, 2006; Van Berkum, Hagoort, & Brown, 1999). This component has been seen to occur when the vehicle cannot be applied literally to the topic (*A* in the same sentence) (Pynte, Besson, Robichon, & Poli, 1996; Blasko & Kazmerski, 2006). The observations reported above therefore suggest a greater effort is needed for the integration of semantic elements in metaphorical sentences than in literal sentences. Sentences containing puns, and punchlines of jokes have also been shown to elicit an N400 (Coulson & Van Petten, 2002; Ritchie, 2005; Tartter et al., 2002), lending further support to the idea that these forms of humour rely on unexpected events to destabilise and reorient the semantic picture built in an utterance (Blasko & Kazmerski, 2006; Coulson & Van Petten, 2007).

A similar pattern has been observed in response to incongruous chords in music (Hantz et al., 1997; Koelsch, Gunter, Schroger, & Friederici, 2003). Further studies have demonstrated an effect of priming music on the amplitude of the N400 associated with word targets (Koelsch et al., 2004; Steinbeis & Koelsch, 2008b) and an effect of priming words on the amplitude of the N400 associated with musical targets (Daltrozzo & Schön, 2009a, 2009b). These sets of evidence, combined with a strong reliance on the semantic interpretation assumption of the N400 component has led some to argue for semantic processing in music — the notion that musical elements can communicate specific concepts (e.g. Daltrozzo and Schön, 2009, Koelsch et al., 2004). Hantz et al. (1997) found a large N400 component in response to in-key (diatonic) but closure-preventing chords placed at the end of harmonic sequences, such as a dominant instead of a tonic. Out-of-key chords (chromatic) produced a shorter negativity followed by a late positivity. These led to the conclusion that the N400 was a product of a lack of closure rather than a reflection of the key-based expectedness of chords. Since incongruous chords have typically been placed either at the end of stimuli or in positions in the stimuli which prevent harmonic closure (e.g. interrupted and unresolved cadences in Steinbeis, Koelsch and Sloboda's (2006) study), the semantic interpretation of the odd chord is difficult to disentangle from the closure interpretation of these effects.

A seminal study in the area of music and language ERP research demonstrated what the authors reported as an identical component in music and language following the processing of unexpected syntactic elements (disguised embedded relative clauses) and out of key chords (Patel et al., 1998). Contrary to the types of musical sequences used in studies reporting a late negativity, these studies used sequences in which the incongruous chord did not prevent the closure and in-key resolution of the musical sequence. The component, a positive deflection 600ms after the onset of a stimulus (P600) is known to be elicited following the processing of syntactic incongruities in language (Osterhout et al., 2002). Its potential relevance to understanding affective responses in both domains was highlighted in a recent study which found a P600 after a functional shift(Thierry et al., 2008). This figure of speech, used for dramatic effect since the days of Shakespeare (Thierry et al., 2008) consists of the use of a noun in the place of a verb (e.g. *The singer was so bad the crowd started to* tomato *him*). Linguists have argued that the lack of linguistic transparency in a functional shift enhances meaning by forcing the listener or reader to resort to context and common-ground knowledge for the interpretation of the utterance (Clark & Clark, 1979). Though the empirical study of the role of this figure of speech in eliciting affective responses is still in its infancy, the idea of extra integration efforts claimed by Clark and Clark (1979) seems promising, in view of the previously mentioned relationship between degree of fit and the amplitude of ERP components. Alongside the investigation of the role of incongruities in eliciting aesthetic-affective responses, this thesis will attempt to reconcile the studies showing a late negative component with those showing a late positive component in music to arrive at a theoretically and empirically satisfactory interpretation of these two distinct patterns patterns.

1.1.4 Are enough similarities present between music and language to warrant a shared model of aesthetic experiences?

Music and language clearly differ in a number of ways. Language refers to extra-linguistic elements while music is self-referential (Besson & Schon, 2001). This means that music creates meaning by varying the relationships between its own elements, by playing with tension and resolution (Patel et al., 1998), while linguistic elements, from the level of the morpheme upwards, refer directly to specific elements and concepts without the need of a wider linguistic context. Though the wider linguistic context does help refine meaning (e.g. the meaning of *bank* in the context of a discussion about rivers or about money), this difference affords much more immediate and more specific meaning to language than to music (Patel et al., 1998). Furthermore, syntactic constraints are much stronger in language than in music (Bod, 2002), with syntactic violations compromising the comprehensibility of sentences in language but rule violations contributing to extra meaning in music (Patel et al., 1998; Lim, 2006). Once more, a nuance must be added in the fact that figurative language often does violate rules of syntax for the sake of emphasis (e.g. the crowd started to *tomato* the singer), but for the most part, meaning is achieved in language by abiding by syntactic rules. This distinction between the two domains is explained by the fact that linguistic dependencies are obligatory while musical dependencies are probabilistic (Patel et al., 1998). Another difference noted by Patel (2003) lies in the lack of an equivalent for linguistic grammatical categories attached to the lexical entry itself in the music domain (e.g. the verb *to be* will always be a verb no matter what sentence it is used in, whereas the note *C* can be a tonic, a dominant, a leading note or hold a number of other musical functions depending on the key of the piece in which it is used). A reversed lack of equivalence lies in the fact that music is understood on the basis of foreground/background relationships (Vuust, Roepstorff, Wallentin, Mouridsen, & Ostergaard, 2006) — a phenomenon absent in language.

Despite these differences, a rapid exploration of theoretical and empirical work carried out across both domains demonstrates a significant number of parallels between music and language. These stem from the fact that both music and language are infinitely combinatorial (McMullen & Saffran, 2004), consisting of a limited pool of perceptually discrete elements which unfold over time and are organised in rule-governed hierarchically structured sequences (Besson, 1999; Brown, 2001; Jentschke et al., 2005; McMullen & Saffran, 2004; Patel, 2003, 2005; Raffman, 1993; Tettamanti & Weniger, 2006), which have larger meanings than the sum of their elements (Brown, Martinez, & Parsons, 2006).

In both domains, perception is facilitated by prior knowledge (Sundberg & Lindblom, 1976), and the perceptual input undergoes a process of hierarchical structuring which is not found in the input itself (Bod, 2002). The bottom-up perception of stimuli is facilitated in both domains by the top-down influence of memory (Jurasfky, 2003; Saffran, 2003) and implicit knowledge about the rules governing each domain (Bod, 2002; Sundberg & Lindblom, 1976). Various models have been used to understand this top-down facilitation. The simplicity principle which states that the simplest interpretation is the one adopted by listeners has also been applied in both domains (Bod, 2002; Collard, Vos, & Leewenberg, 1981; Frazier, 1978; Lerdahl & Jackendoff, 1983). The likelihood principle, which states that ambiguous stimuli are interpreted according to the most likely interpretation based on statistical probabilities within each domain, has also been used to model perception of both music (Bod, 2001; Raphael, 1999) and language (MacDonald, Pearlmutter, & Seidenberg, 1994; Manning & Schütze, 1999), with a combination of both principles proving most successful across both domains (Bod, 2002). In both music and language, disfluencies and errors are substituted with correct readings and hearings (Longuet-Higgins, 1994), and plausible substitutes are produced during imperfect recall (Patel, 2003). Thus, although music and language are physically different, strong similarities are seen in the computations involved in their perception (Besson, 1999), and overlaps have been demonstrated between brain areas involved in processing both domains.

Central to this thesis, both music and language generate strong expectations (Besson & Schon, 2001) and their processing necessitates keeping track of structural dependencies while performing the structural integration of incoming discrete elements (Tettamanti & Weniger, 2006). In both domains, top-down-facilitated automated and unconscious processes (Swinney & Love, 1998) rely on the internalisation of regularities learned in an incidental manner through passive exposure (Jentschke et al., 2005). The integration cost of each incrementally encountered element has been associated with expectations in both domains. The greater the distance between the predicted element and the actual element, the greater the processing difficulty (Patel, 2003), and the greater the number of possible conformations that need to be discarded from the set of all possible conformations as each new piece of information is processed, the greater the computational cost (Tettamanti & Weniger, 2006).

Kutas (1998) equated differences in ERP patterns, associated with computational costs, to differences in mental activity; this has led some researchers to assume that identical electrical patterns in response to two different stimuli indicate identical mental activity. Though not uncontroversial (Gaillard, 1988), the widespread adoption of assumptions such as these in the interpretation of ERP effects has led to the conclusion that similar processes underlie the detection and integration of unexpected elements in both music and language. Patel et al. (1998) found statistically indistinguishable patterns in response to syntactically improbable elements in language (embedded subordinate clauses) and harmonically incongruous chords in music (chords from a different key to the key of the piece). This observation, amongst others, led to the suggestion of shared syntactic resources between music and language (Patel, 2003). In a more controversial study, Koelsch et al. (2004) demonstrated a statistically indistinguishable N400 in response to words in a priming task in which the primes were either sentences or musical excerpts, suggesting that music could act as a semantic prime, while acknowledging a difference in the nature of semantic information conveyed by music and language.

However compelling these data may seem, with methodological development comes the responsibility of clearly defining what the data signify for psychological theory, and many questions as to both the mechanisms underlying the elicitation of the ERP patterns described above and their role in eliciting affective responses remain unanswered (Steinbeis, Koelsch, & Sloboda, 2006; Thierry et al., 2008). Indeed, while the literature supporting the similarities in neural activity between music and language processing proliferates (Besson & Macar, 1987; Patel et al., 1998; Patel, 2003; Koelsch et al., 2004; Koelsch, Gunter, Wittfoth, & Sammler, 2005) it is for the most part devoid of the empirical investigation of the role of incongruities in the affective-aesthetic appreciation of music and language. Furthermore, studies aiming to establish a role of incongruities in affective aesthetic responses are being conducted separately in music psychology (Steinbeis et al., 2006; Vuust et al., 2006) and language psychology (Coulson & Van Petten, 2002, 2007; Thierry et al., 2008).

This state of affairs leaves authors across these domains in danger of leaps of faith as to both the existence of shared mechanisms underlying the detection and integration of incongruities in both domains and the role of these incongruities in eliciting affective aesthetic responses.

1.2 Conclusion

The research carried out during this PhD, based on these observations, had the aims of providing a means of testing theories of incongruity-based aesthetics in music and language, empirically testing these theories and investigating the cognitive and neurophysiological underpinnings of these heightened aesthetic experiences. Alongside these primary goals, the methods adopted allowed an explanation of discrepancies between studies in this field, insights into the interpretation of ERP components, a greater understanding of the processing of rhythmic incongruities, and discoveries of significant effects of musical training on both music and language processing.

Kutas' (1998) review of language structures concluded by stating that the study of language cognition was useful in understanding structures underlying general cognitive principles of memory access and knowledge organisation. Since music and language both involve structures which unfold in time and are both processed according to systematicities which are for some universal (e.g. the existence of phonemes and notes) and for others culture-specific (e.g. the specific sets of phonemes and notes and their combination rules, McMullen and Saffran, 2004), I would argue that the same statement applies to the study of music cognition, and to a greater extent, to the parallel study of music and language cognition, making the findings of this research of benefit outwith the fields of music and language psychology.

Chapter 2

Reviewing aesthetic incongruities: a case for the role of closure in affective responses to music and language

2.1 Introduction

The temporary disruption in the construction of a stable representation of an incrementally processed stimulus has long been suggested to underpin aesthetic and affective responses elicited by both music and language. Researchers across both domains recognise the need for more systematic empirical research into incongruity-based models of music and language aesthetics (Juslin & Västfjäll, 2008; Thierry et al., 2008). Furthermore, though incongruities seem to contribute to aesthetic-affective responses in both music and language, the processing of aesthetically-purposed incongruities in both domains have not been systematically studied side by side.

However, before any such research can be carried out, a systematic review of the literature is needed, to establish whether the empirical work carried out to date in music and language psychology supports the plausibility of a shared mechanism of aesthetics across these two domains of human cognition. This chapter presents a systematic review of the literature published to date with the aim of answering the following research question: Could the same mechanism, underpinned by the detection and integration of incongruities, underlie the eliciting of aesthetic-affective responses to music and language?

2.2 Addressing the research question

2.2.1 Three steps to an affirmative answer

Before delving into the literature, it is necessary to establish a set of conditions which would need to be met in order for the answer to be affirmative. Firstly, the literature would need to show evidence of similar mechanisms in the perception, encoding and retrieval of patterns formed by diverse elements. Secondly, studies would need to demonstrate similarities in the mechanisms resulting in expectancies and in the processing of incongruous elements which violate those expectancies. Thirdly, similarities in affective responses following the detection and integration of incongruities and in the absence of other sources of emotions (such as overall meaning or musical connotations) would need to be evidenced. Finally, the literature used to fulfill these three conditions must stand up to the scrutiny of psychological, computational and biological plausibility.

2.2.2 Key terms

Another prerequisite to the exploration of the literature is the clear definition of the key terms of the research question.

2.2.2.1 Music

In his later work, Meyer (1974) makes reference to the hierarchical nature of music, whereby "tones [notes] combine to form motives, motives cohere forming phrases, and so on, until the highest structural level (that of the movement or work as a whole) is reached" (p. 51). He goes on to suggest that the nature of the structures which bind these elements together should form the basis of scientific enquiry. This review will consider of relevance to the research question publications which study the processing of music from the level of the note upwards. Unless stated otherwise, the term "music" refers to music without lyrics. No studies concerned with the perception, encoding and retrieval of components of a musical note (e.g., combined frequencies, spectrum analysis) will be considered. Furthermore, it is the structure of the whole into which the notes, chords, phrases and so on are organised which is of interest. Its production, texture, aesthetic execution and wider contextual setting have formed the basis of much psychological enquiry to date, and lie beyond the scope of this review. Consequently, studies concerned with timbre, dynamics, and tempo were excluded, in favour of studies investigating the processing of elements tied to the deeper structural aspects of the music itself, which would remain constant were the music played on a different instrument or in a different extra-musical context.

2.2.2.2 Language

In a similar vein, this review is concerned with the strictly verbal-linguistic aspects of language (Ross, 2010), namely the processing of semantics and syntax. Equating the note with the smallest unit of meaning in language, the morpheme, one could say that morphemes combine to form words, words cohere forming sentences and so on until the highest structural level (that of the discourse or work as a whole) is reached. I therefore excluded publications concerned with aspects of language processing below the level of a morpheme (e.g., formants, spectral analysis). The exclusion of music studies concerned with timbre, dynamics and tempo was paralleled here with the exclusion of studies concerned with paralinguistic elements and wider semantic context. While recognising their importance in conveying emotions in language, these factors must be excluded here in order to allow both a clear focus and a parallel approach to the study of how the structural juxtaposition of elements of music and language might contribute to aesthetic affective experiences in both domains.

2.2.2.3 Affective and aesthetic experience

In refining the notion of affective experience, a first distinction must be made between the portrayal and the eliciting of affective states. This distinction is respectively between what a listener believes is being depicted in the music and the language, and what he or she experiences. It is the latter — the eliciting of affective states — which forms the object of this enquiry, the "immediate aesthetic delight " (Aiken, 1951, p. 29) experienced in the perception of a work of art. A second distinction lies in the nature of this experience. Meyer (1956) distinguishes affect from emotion, defining affect as the "feeling-tone which accompanies emotional experience" (p. 12). Mandler's (1982) approach, endorsed by Ortony, Clore, and Collins (1988) distinguishes between "cold" (cognitive interpretation) and "heat" (arousal resulting from the unexpected). The term aesthetic-affective response is used throughout this thesis to refer to the "aesthetic delight" and "heat" component of participants' responses to music and language, such as thrill, the feeling of being moved, and enjoyment of the stimulus, and, on the other end of the spectrum, boredom, annoyance and disinterest (Brattico & Jacobsen, 2009). As such, it covers both sensations of aesthetic pleasantness and affective responses to the stimulus. These would arguably benefit from being distinguished from each other. Experiment 3 in this thesis provided a means of testing these separately from each other, but until the presentation of that analysis (Chapter 7), this distinction is difficult to achieve based on previous literature published in this area of research.

2.2.2.4 Literature

The literature examined here was restricted to peer reviewed publications which presented original data of empirical studies in music and language psychology. These publications included those presenting experiments jointly investigating music and language, or reporting data from one domain within a strong theoretical understanding of the other domain (e.g., music psychology data analysed within the frame of linguistic cohort theory, Bella, Peretz, & Aronoff, 2003).

2.2.3 Databases

Since the review was concerned with the psychological literature surrounding language and music cognition, searches were conducted via the *Ovid* interface with the *PsycInfo* database, through the studies listed from 1801 to present. Since the literature to be included in the review potentially also included publications from linguistics and music studies, the identical searches (see below for key terms) were carried out in *Web of Science*. Other, less specifically experimental databases such as *JStor* were not included, as the experimental literature contained in such databases also featured in the *Web of Science* database. Further limitations on the searches included the restriction to papers published in English or in French, since from a scoping exercise it was apparent that publications relevant to the research question either originated from French-speaking or English-speaking countries or were published in English. A final restriction imposed on the search level of this method was the necessity for studies to have been carried out on human participants. Experimental studies which were not found through this search strategy but were mentioned in found studies were manually retrieved and analysed. Special issues such as the New York Academy of Sciences' *Biological foundations of music* and special issues of the *Revue de neuropsychologie* were hand searched.

2.2.4 Search terms

The search terms used in this systematic research review were developed in several stages. In order to scope the literature, initial searches were carried out in *Web of Science* alone using the search terms 'music *AND* language', leading to the retrieval of 1110 articles published since 1900. Through reading the abstracts, 229 were retained as potentially relevant to the questions to be addressed in this research.

The key words in the relevant studies were used to develop the terms music and language with other terms related to those domains (e.g. melody, rhythm, sentences, lexical). The keywords from the studies deemed not relevant to this area of research (mostly those focussing on cultural or social aspects of music and language) were also noted. These were used as exclusion terms if they were found in more than four such articles retrieved during the scoping exercise. So as to ensure that no exclusion search terms would compromise the successful retrieval of all relevant articles, these were cross referenced with the keywords contained in articles deemed relevant.

Once the keywords from all relevant studies had been cross-referenced with those for non relevant studies, the following search terms were used in *Web of Science* (from 1900 to present) and *PsycInfo* (from 1801 to present), searching through the title, abstract and keywords of each listed article in both databases within the limitations detailed above.

The final set of search terms used in this SRR was as follows:

> (language* OR lexical* OR lexicon* OR linguistic* OR prosodic* OR prosody* OR semantic* OR sentence* OR speech OR spoken OR speak* OR speech OR syntactic* OR syntax OR verbal* OR word* OR write OR written OR writing*)
>
> *AND*
>
> (melodi* OR melody OR music* OR rhythm* OR song* or tempo OR tonalit*)

NOT

(cultur* OR africa* OR identity OR social OR popular OR ethnic* OR adolescent* OR national* OR media OR curriculum OR hip hop OR american* OR health OR nurs*)

2.2.5 Search outcomes

In applying these search terms to both *Web of Science* and *PsycInfo* a total of 4951 references were retrieved. Once duplicates had been excluded, 3973 articles remained.

In view of the large number of hits, the possibility of excluding studies before a specific date was explored. To this end, a shift either in paradigms or in numbers of publications was sought by slicing the studies into decades. No significant paradigm or number shifts were found, invalidating this method of limiting the number of studies included in this review. The abstracts of all 3973 articles were therefore examined to judge the relevance of their research to the aims of this systematic review and articles were included or excluded on the basis of the criteria detailed below.

2.2.6 Inclusion and exclusion criteria

As stated above, studies were only retained if they investigated either both the cognition of music and language, or the cognition of one in the light of the other. However, these criteria alone were not sufficient to narrow down the studies to those directly relevant to answering the research question. A systematic approach to limiting the retained studies to those addressing the areas of music and language cognition detailed in the introduction was needed. To this end, the following inclusion and exclusion criteria were developed. In order to be retained for final analysis in the present systematic review, studies had to:

1. use music and/or language as the stimulus of interest rather than as a distractor from the main task of an experiment;

2. either experimentally investigate both music and language cognition or experimentally investigate one within a strong theoretical background of the other; studies investigating only music or only language were excluded from the analysis;

3. report new experimental data; reviews of sets of experimental studies and theoretical accounts were excluded from the analysis but some were retained for narrative purposes;

4. have a strong theoretical background; studies which simply collected brain imaging data from language and music or mention language or music only after having identified brain areas known to be involved in one or the other domain were excluded from the analysis but taken into consideration to reinforce the psychobiological plausibility of the outcomes of this analysis;

5. use only healthy human participants; studies involving only cognitively or perceptually impaired populations were excluded;

6. investigate the deep cognition of music and language; this criterion was developed based on the notion of levels of processing developed by Craik and Lockhart (1972); within this approach, two levels of processing can be distinguished: the processing of the physical or sensory features of a stimulus and the matching of an input against stored abstractions from past learning which leads to pattern processing and the extraction of meaning; studies concerned with the processing of acoustic characteristics of musical and linguistic events, rather than with the processing of the structure of music and language were excluded;

7. investigate the processing of music and language rather than the processing of components of musical and linguistic events; studies investigating the processing of elements below the level of the smallest meaningful unit in language (morpheme) and music (note) were excluded;

Of the 3973 distinct articles retrieved, 294 were judged experimental from the reading of their abstract. The exclusion criteria detailed above were applied to these 294 studies, bringing the final number of studies retained for full retrieval, further analysis and potential inclusion in the research review down to 81. Following retrieval, a further 25 studies were excluded because they failed to meet the inclusion criteria or presented previously published data. This selection process resulted in the retention of 56 articles to be reviewed systematically.

A final search, using the methods described above, was carried out during the summer of 2010 to retrieve any more recently published articles fit for inclusion in the systematic review. This search yielded eight further articles fit for inclusion. Articles excluded from the systematic review which were relevant to the research question were retained as additional discussion material.

2.3 Step 1: Does the literature show similar patterns in the perception, encoding and retrieval of music and language?

Some similarities between music and language cognition, such as primacy, recency and suffix effects (Greene & Samuel, 1986; Roberts, 1986) are common to the processing of other types of stimuli, such as symbols (Jarvik, 1951), abstract visual arrays (Broadbent, 1981) and faces (Smyth, Hay, Hitch, & Horton, 2005). Other areas investigated in both music and language are motor-based perception (Callan et al., 2006) and the role of attention in heightening sensitivity to stimuli (Segalowitz & Plantery, 1985). Though of interest in determining common patterns of general cognition, the present review focused solely on aspects of cognition which are specific to music and language. This specificity is often centred around the way in which both are considered combinatorial, meaning that larger structures are generated hierarchically from a pool of smaller, more unitary components (Brown et al., 2006). It is the commonality between the processes underlying this perceptual structuring, central to the creation of tension and resolution and therefore aesthetics and affect (Meyer, 1956; Ortony, 1975), which is under investigation in this first part of the review.

2.3.1 Models applied across both domains

The cohort model of lexical access (Marslen-Wilson & Tyler, 1980) states that word recognition is achieved through the successive reduction in the number of possible word candidates as each new phoneme of an incrementally processed word is perceived. According to this model, candidate words are gradually eliminated from the cohort of possible words as the listener incrementally processes more information about the word. It is only when the cohort of possible words is narrowed down to one remaining candidate that successful word recognition is achieved. This phenomenon has repeatedly been demonstrated in the recognition of auditorily presented words using a gating paradigm (Van Petten, Coulson, Rubin, Plante, & Parks, 1999). Interestingly, when Van Petten et al. (1999) investigated the relevance of the cohort model to the recognition of words within a sentence, the authors noted that semantic integration was not only tied to word identification but also to a sentence's semantic context, in that word meaning appeared to be processed before the listener had been given sufficient acoustic information to uniquely identify a word. This same paradigm was applied to the recognition of melodies in a study by Bella, Peretz, and Aronoff (2003), in which participants were presented with one note, then the first two notes, then the first three and so on until the tune was recognised. The patterns in the data were interpreted within the cohort model: the fact that familiarity diminished recognition time was attributed to the fact that more familiar melodies will have a higher initial activation within a cohort, due to their frequency in the listener's past experience. These observations bring into play not only how the listener compares the candidates in the cohort to the incrementally processed incoming sounds, but also how the cohort of potential candidates is organised, through exposure, in the listener's mind.

In a study published the same year, Schulkind, Posner, and Rubin (2003) investigated the time course of melody recognition adding to a general cohort model the notion of structure within a piece. The analysis of the point at which melodies were identified revealed that notes upon which participants recognised famous melodies were most often located at phrase boundaries, notes denoting the end of a particular sequence and notes which fell on accented beats. The effect of the structural role of notes on their impact on melody recognition suggests that structure-facilitated processing, highlighting points of interest within a piece, is reflected in the way melodies are stored in the brain. This evidence for the hierarchic structuring of information echoes previous studies demonstrating a migration of the time localisation of noise bursts towards phrase boundaries in participants' recall of continuous streams of language (Fodor & Bever, 1965) and music (Sloboda & Gregory, 1980). Building upon these findings, and upon Jarvella's (1970) finding that recall of words within connected discourse was best when the words were within syntactic phrase boundaries, Chiappe and Schmuckler (1997) conducted a music study in which participants were asked to indicate whether a seven-note probe matched the last seven notes of an unfamiliar melody to which they had just listened. The probe either fell within or across a phrase boundary. Since probes contained within phrase boundaries proved to be better recall cues than those crossing boundaries for musically trained participants, Chiappe and Schmuckler (1997) concluded that "musical phrases are distinct functional units during musical processing" (p. 257).

2.3.2 Insights from neuroimaging

These two sets of examples highlight commonalities in the way incoming streams of information are incrementally organised into perceptually stable hierarchical structures in the mind of the listener. The chunking of information into distinct hierarchical units, by mapping input onto schema acquired through exposure to a finite probabilistic grammar (Tillmann et al., 2000), appears to be common across both domains. Results from neuroimaging studies also support similarities in the processing of both types of stimuli, and point to the possibility of a common "structure tracker" for music and language (Levitin & Menon, 2005). Other similarities between the neural areas activated during music and language processing were detailed by Mirz et al. (1999).

The Primary Auditory Cortex (PAC, Brodmann's Area (BA) 41/42) has long been established as the starting point for the processing of auditory stimuli (see Celesia, 1976, for an early review of the literature). From this area, activation patterns spread toward the superior temporal gyrus (BA 22) (Mirz et al., 1999). The observation of a bilateral activation in BA 22 during the processing of musical and verbal stimuli, both heard and imagined (Ducreux, Marsot-Dupuch, Lasjaunias, Oppenheim, & Fredy, 2003; Schmithorst, 2005), supports this observation. Brown (2006) found evidence that BA 44 and BA 45, which together form the area referred to as *Broca's area*, and right hemisphere homologous areas, were activated during both the processing of complex stimuli in music and language, where complexity is defined as a function of an increasingly structured nature of the elements which the listener integrates to form a meaningful whole (e.g. with increasing complexity: tones, melodies, harmonised melodies, orchestral music; morphemes, words, sentences, discourse). The authors suggested that this activation could reflect the participants making use of implicit knowledge of harmony and syntax, with BA 44/45 acting as an interface between syntactic functioning and phonological generativity as suggested by Hagoort (2005) who added to this picture the processing of semantics in BA 47. Brown et al. (2006) used these imaging findings to suggest that the processes involved in sequential ordering in music may be homologous to those underpinning language syntax. Though the activity in these areas could merely be reflecting a heightening of attention due to the complexity of auditory stimuli, the repeated observation of activation in BA 47 in the processing of structured visual stimuli in mathematics (Martin-Loeches, Casado, Gonzalo, De Heras, & Fernandez-Frias, 2006) reduces the plausibility of an auditory-stimulation artefact and supports the suggestion that this area has a role in the processing of structure and syntax in both visually and auditorily presented stimuli outwith language.

Using Magnetoencephalography (MEG) to localise the source of the electric activity patterns seen at the scalp in an ERP study by Koelsch, Gunter, Friederici, and Schröger (2000), Maess et al. (2001) demonstrated that neural patterns signaling the processing of syntactic incongruities in music had their source in BA 44/45, paving the way for the investigation of whether music syntax is processed in the same areas as language syntax. More recently, Levitin and Menon (2005) found a significantly higher activation in BA 47 in the left hemisphere when non musicians listened to normal music compared to when they listened to scrambled music, which was devoid of both harmonic and rhythmic structure. The authors concluded that this area, adjacent to Broca's area (BA 44/45) which has been widely recognised as playing a key role in the processing of syntactic structure in language, was involved in the processing of structure in music. These findings replicate those of Ducreux et al. (2003) who investigated which areas showed activity in the absence of any auditory stimuli when participants were asked to imagine the melody, lyrics or both of a song. This functional Magnetic Resonance Imaging (fMRI) study also revealed activity in BA 22, 45 and 47 as well as in the hippocampus, for both music and language, suggesting that these areas are active in the processing of both domains in the absence of external auditory input. This in turn negates claims that these areas are merely activated in response to complex auditory stimuli. However, the extent to which participants would have been able to think of the melody of a well-known song without thinking of its associated lyrics was later questioned by Miranda and Ullman (2007), citing evidence for an integrated memory for the lyrics and melody of famous songs (Peretz, Radeau, & Arguin, 2004).

A recent study using fMRI investigated the integrated vs. segregated processing of lyrics and music in songs (Sammler et al., 2010). Studies have shown that repetition of a stimulus lessens the activation of areas involved in processing the stimulus. This paradigm was used to assess whether music and lyrics were processed in an integrated fashion or separately, by varying either the lyrics or the melody of unfamiliar tunes.

Sammler et al. (2010) demonstrated that the extent of integration varied along a posterior to anterior axis of the left superior temporal sulcus and left precentral gyrus, showing an increased degree of integration posteriorly compared to anteriorly as processing progressed from a more basic acoustic level to a higher conceptual level.

Groussard et al. (2010) used PET scans to investigate which areas of the brain showed greatest activation when participants were asked to remember lyric-less tunes and words which they had knowledge of without knowing when or where they had acquired that knowledge. The authors referred to this as semantic memories. The data pointed to a common network throughout the temporal cortex, with an anterior preference for musical memory and a posterior preference for verbal memory, showing a degree of overlap between the areas involved in retrieving music and language. Since the medial temporal lobe is known to be involved in memory (Squire & Zola-Morgan, 1991), and since the verbal instructions would have been recalled during memory for music, it could be argued that this overlap was due to the identical retrieval task performed by the participants, rather than to the way the musical and semantic lexica are organised in the brain.

Imaging data alone cannot inform cognitive processes (Uttal, 2001). Nonetheless, these patterns support the biological plausibility of a structure-based processing system in music and language, involving some of the same neural areas in both domains. Furthermore, theorising the involvement of neural areas traditionally associated with the processing of language in the online structuring of musical input elucidates discrepancies observed in early brain localisation studies.

Studies concerned with language prosody and paralinguistic cues such as gestures and facial expressions have reported the involvement of the right hemisphere (see Ross, 2010, for a review of brain localisation studies in language). However, when considering only the strictly verbal-linguistic studies as described at the outset of this review, an early consensus in the joint study of music and language psychology placed the processing of language in the left hemisphere and that of music in the right hemisphere. This consensus stems from the observation of ear preferences in the recognition of musical or verbal material (Kimura, 1964; Goodglass & Calderon, 1977; Segalowitz & Plantery, 1985) and task interferences using delayed auditory feedback (Bradshaw, Nettleton, & Geffen, 1971). This early consensus was however challenged by subsequent studies. An early ERP study showed activation patterns in the left hemisphere during language processing and in both hemispheres during non-lyrical music processing (Konovalov & Otmakhova, 1984). Both Konovalov and Otmakhova (1984) and DePascalis, Marucci, Penna, Labbrozzi, and al. (1987) found that musical training reduced the hemispheric asymmetry of neural activity linked to the processing of music while the left-heavy processing of language remained unchanged. Konovalov and Otmakhova (1984) claimed that this was due to different aspects of music being processed in different hemispheres while DePascalis et al. (1987) claimed this pattern reflected less difficulty for musicians in identifying and chunking musical themes. Both interpretations are consistent with the previously mentioned evidence supporting more elaborate cognitive templates in musicians on to which incoming music can be mapped and a greater degree of structure processing supported by the left hemisphere.

LaBarba and Kingsberg (1990) demonstrated that participants preformed less well in a right-handed task (reliant on the left hemisphere) when listening to orchestral music. This study suggests the involvement of left-hemisphere resources in orchestral music processing. Furthermore, Cheatham (1990) failed to replicate a left-ear preference for music processing when using orchestral music, finding no significant difference between ear preference for music and language. Cheatham (1990) argued for a general preference for right ear processing in right-handers. Neither of these studies challenge the notion that language is predominantly processed in the left hemisphere; instead, they challenge the restriction of music processing to the right hemisphere. To account for apparent disparities between studies, Mirz et al. (1999) suggested that differences in activation patterns could be due more to stimulus complexity (digits and tones vs. sentences and melodies) than to stimulus type (music vs. language). Indeed, careful study of the stimuli used in these studies reveals that on the whole those reporting the involvement of the left hemisphere in music processing used orchestral music (Cheatham, 1990; DePascalis et al., 1987; Konovalov & Otmakhova, 1984; LaBarba & Kingsberg, 1990), while those showing right hemispheric specificity used less complex stimuli such as three-tone-long sequences (Goodglass & Calderon, 1977) and unharmonised melodies (Kimura, 1964). These data patterns suggest once more the involvement of the left hemisphere in the processing of only music and language which require the listener's mind to organise elements into a meaningful complex structure (e.g. meaningful sentences or music extracts rather than digits or tones).

2.3.3 Similar processes, overlapping areas, but shared resources?

This distinction between complex and simple stimuli can also be used to account for discrepancies between studies investigating the shared nature of resources used for the processing of music and language. Behavioural evidence from dual-task interventions includes a study by Bradshaw et al. (1971) in which a domain-specific interaction under delayed auditory feedback was found. This observation suggests that music and language are processed independently from each other, since only delayed auditory feedback in the domain being used in the task affected performance. The relative independence of music and language encoding for later retrieval was investigated by Goodglass and Calderon (1977). Since the memory patterns in the dual task were identical to those observed when both types of stimuli were listened to separately, the authors concluded that "independent parallel processing takes place in the two hemispheres for their preferred (verbal vs. tonal) components of a complex stimulus" (p. 403). However, the stimuli referred to as "music" were formed of three tones, with either one low tone and two high tones or vice versa, which is not subject to hierarchical structuring. Thus, it could be that these stimuli do not call upon common resources because of the lack of any necessity to integrate the elements into a meaningful structure.

The findings that rhythmic information in music is retained in the phonological loop

(Saito & Ishio, 1998; Schendel & Palmer, 2007) and that language tasks, be they the articulation of vowels (Saito & Ishio, 1998; Schendel & Palmer, 2007) or non articulatory verbal tasks (Saito, 2001), interfere with short term memory for musical rhythms also suggest a shared resource between the two domains. The fact that white noise does not interfere with memory for music or language (Schendel & Palmer, 2007) strongly suggests that the two domains share an encoding and retrieval mechanism which is separate from the storage of other types of auditory information. These data are also consistent with the notion of resources being shared only when integration is necessary, as rhythmic stimuli require temporal integration, which can be defined as the combination of perceptually discrete elements into structured sequences governed by previously acquired knowledge of regularities (Jentschke & Koelsch, 2009). More recent studies investigating the phonological loop's role in memory for music have demonstrated an interference of letter memory on memory for musical stimuli (Williamson, Baddeley, & Hitch, 2010). However, as in early studies, the "musical" stimuli were a set of three tones which did not constitute a tune and which were labeled as high, middle or low. Furthermore, since recall was indicated via a response grid on which participants ticked boxes in the high, middle or low row, it is quite possible that the labeling of these stimuli in the instructions led to verbal encoding of the stimuli, which would account for the observed interference. Studies using stimuli such as these do not fit the criteria for inclusion in this review. However their findings help clarify this argument and were included for that reason.

Defining syntax as "a set of principles governing the combination of discrete structural elements (such as words and tones) into sequences" (p. 674), Patel (2003) set forward the Shared Syntactic Integration Resource Hyphothesis (SSIRH). Following the observation that linguistic and musical syntactic integration rely on common processing regions, this model, which assumes limited syntactic processing resources, states that tasks which combine linguistic and musical syntactic integration should show a degree of interference between the two. This model, recently revisited by the author (Patel, 2007), captures the data trends observed above and remains the most theoretically and empirically robust model to date. It also serves to explain the fact that the levels of activation of BA 47 were unaffected by musical training (Levitin & Menon, 2005) while the hemispheric asymmetry for areas supporting other aspects of music and language processing were: if one considers the evidence for more intricate mental representations of music (Deliege, 1996) and more top-down influence of mental representations on on-line parsing (Chiappe & Schmuckler, 1997; Schulkind et al., 2003) in musically trained participants, the differences in asymmetries between trained and untrained participants can be explained by a greater use of music-specific mental representations stored in the right hemisphere in musicians during the encoding of musical stimuli. These stimuli are strategically listened to according to the top-down influence of these representations, thus relying less on integration processes supported by the left hemisphere. In the case of nonmusicians, the structure tracker must work harder to integrate less complex representations at a lower level of the hierarchic structure into a greater whole, relying on the left hemisphere more for bottom-up encoding and calling upon the right hemisphere less to supply top-down facilitation.

Participants' responses to elements which pose a problem to this syntactic integration are discussed in Step 2 of this review. However, the framework of this theory is useful for explaining the literature dealing with the extent to which resources are shared. The fact that Sammler et al. (2010) demonstrated more segregated rather than more integrated processing as complexity and cognitive engagement increased seems initially to contradict this claim. For several reasons this is not the case. Firstly, the conceptual processing to which the authors refer is more closely tied to semantics than to the structural processing of the music. In the examples given, the syntactic structure of the sentence was kept very similar across the different lyrics associated with the same melody. Secondly, the study supports the notion that the level at which the music or language is processed varies in different brain areas in a passive listening task. It is therefore plausible that different tasks, tapping into different neural resources and processes, would generate a different amount of interference between the processing of music and language stimuli. The notion that simple unstructured stimuli can be processed separately but that the structural integration of distinct musical or linguistic elements is reliant on shared resources is thus not challenged by those findings. Indeed, it is supported by literature explored in section 2.4. This framework shows a good fit with the data gathered in the literature dealing with the extent to which cognitive processes and mental resources are shared across these domains.

2.3.4 Step 1 — Summary

The observations above suggest strong similarities in the processing of music and language stimuli, notably in the importance of organising the perceptual input into stable hierarchically organised structures. Furthermore, the research reviewed supports the idea that implicit knowledge in both domains is accessed online and used in a similar fashion to facilitate the processing of stimuli. The question of whether these processes relied upon the same resources was met with conflicting answers. Interference between concurrent music and language tasks only seemed to occur if the tasks involved processes previously linked to the activation of structure-tracking areas. A pattern therefore emerged of shared structure-tracking resources but independence in the processing stages leading to and following the processing of structure in both domains. The centrality of complete and integrated hierarchical units was highlighted in memory studies, and suggests a striving for closure and stability in music and language processing. These conclusions enable us to answer yes to the first step of this research question, namely to support the psychological, computational and neurological plausibility of similarities between music and language cognition, paving the way for a discussion of the literature reporting the similarities in the processing of structure-violating stimuli in both domains — the next step in establishing the plausibility of a common effect of incongruities on aesthetics and affective responses in both domains.

2.4 Step 2: Is there evidence for equivalence between music and language in the processing of incongruous elements?

2.4.1 Defining incongruities

The previous section outlined similarities between music and language in the way stimuli were mapped onto mental templates with what appears to be a facilitatory role of the template-congruence of the incoming stimuli for encoding and later retrieval. This section discusses the similarities and differences seen in the behavioural, cognitive and neurophysiological consequences of template-incongruent elements, described as central to aesthetic and affective responses in both domains (Meyer, 1956; Ortony, 1975; Steinbeis et al., 2006; Thierry et al., 2008).

The understanding that these templates are acquired as a result of repeated exposure (Tillmann et al., 2000) supports their conceptualisation as a set of probabilistic relationships (Krumhansl, 1985; Hale, 2001). An element described as incongruent is one which has a relatively small probability of occurring, either because it is erroneous (e.g. a verb in the wrong tense in a sentence; an out of key note) or because it is not the most probable continuation of a stimulus (e.g. a low-cloze probability word; a chromatic chord). Both psycholinguists and psychomusicologists refer to these elements as unexpected (Van Berkum et al., 2005; Hale, 2001; Levy, 2008; Patel et al., 1998; Koelsch et al., 2004) because of the probabilistic nature of cognitive templates. Although online predictions have been demonstrated in language (Van Berkum et al., 2005), they are as yet only assumed in music. Nonetheless, the facilitatory role of template-congruence demonstrated in section 2.3 in the encoding of musical stimuli suggests that the same applies to this domain.

Incongruities in the studies reviewed here are of two sorts. The first is illustrated in a study by Bonnel, Faita, Peretz, and Besson (2001). The authors asked listeners to detect semantic and melodic incongruities (a nonsensical word; an out of key note) at the end of excerpts of unfamiliar operatic songs. The stimuli presented a combination of semantic-only, melodic-only or simultaneous semantic and melodic incongruities. Considering the high success rate in the detection of incongruities in both musicians and nonmusicians (overall percentage correct trials: instrumentalists 79.5%, singers 80.9%, nonmusicians 76.3%), the data confirm that it is possible to categorise an element of music and language as erroneous in the absence of knowledge of what a specific excerpt should sound like, regardless of musical training. Because these incongruities were detected based solely with reference to the implicit rules of music and language cognitive templates, I shall refer to these as rule-based incongruities.

The second type of incongruity depends on a more specific memory of a particular stimulus. These can be repetition-violating incongruities, as seen in an "oddball effect" such as when an uppercase word appears after a succession of lowercase words (Osterhout & Holcomb, 1992) or a mistuned chord in the midst of a set of the same chord played in tune (Koelsch, Schroger, & Tervaniemi, 1999). These effects, which have been shown to result in a Mismatch Negativity (MMN) or a P300 have been associated with working memory processes (Hantz et al., 1997) and are not specific to music and language processing. Other memory-based incongruities include the processing of stimuli which are modified versions of famous sayings, lyrics or melodies.

The first study to investigate ERP responses to music and language incongruities reported an identical early negativity occurring 100ms after participants heard an incongruous note at the end of a scale or famous tune, and an incongruous (low cloze probability) word at the end of a well-known proverb or novel sentence (Besson & Macar, 1987). Since this pattern occurred both after rule violations and memory violations, it was concluded that it was an indicator incongruity detection, regardless of the reason for the incongruous status of the event. A later study reported a similar pattern at 120ms after the event, in memory-based musical incongruities (Paller, McCarthy, & Wood, 1992). However, these incongruities were notes a tritone away from the expected note of a familiar tune, automatically making them out-of-key notes, thus confounding a memory violation (e.g. not going down one tone at the end of "baa baa black sheep" onto the note sung on the word "lane") with a rule violation (singing the word "lane" on a note which does not belong to the key in which the song is sung).

Another helpful distinction was drawn up by Kutas and Van Petten (1988). Violations of physical parameters, which typically deviate from short term expectancies formed during the experimental sessions and which are not specific to language typically elicit an oddball effect reflected in an early positivity 300ms after the occurrence of the incongruous element (P300). This effect has been reported in response to physically deviant elements (a change in timbre, upper/lower case, tuning and so on; Paller et al., 1992; Hantz et al., 1997; Koelsch et al., 1999; Miranda and Ullman, 2007) and has been associated with working memory processes (Hantz et al., 1997). Contrasted to these are violations of the implicit rules governing the combinatorial hierarchy of a symbol set. These violate the expectancies created by one's long-term memory store concerning the nature of this set of symbols (in the case of language, the assumption that a sentence should make sense and follow a set structure), and elicit components specific to the processing of that set of symbols. Though the behavioural measures and EEG patterns associated with the oddball effect are informative for studying short term memory phenomena, they are not of prime relevance to the overarching research question of whether emotions are elicited in music and language via shared processes since these effects are seen with a wide range of stimulus types. The focus in this section will therefore be on elements which violate the implicit rule governing music and language.

2.4.2 Contributions from EEG data

Because of its temporal precision, the ERP technique has become the most widely used means of investigating the processing of incongruities over the last two decades. If one believes Kutas (1998) about the significance of identical ERP patterns, the evidence from neural imaging studies in music and language strongly supports common processes across both domains.

Following the first demonstration of an N400 ERP component (a negative deflection occurring approximately 400ms after the onset of an element of interest) linked to the onset of a low-cloze probability word (N400) (Kutas & Hillyard, 1980), initial attempts to replicate this pattern in music by presenting participants with music which contained structural incongruities proved unsuccessful (Besson & Macar, 1987; Paller et al., 1992). The first study to attempt this parallel was conducted by Besson and Macar (1987). Participants read sentences which ended with a high or low-cloze probability word (Taylor, 1953) and listened to melodies of famous tunes which ended either with the correct (congruous) note, or with an incongruous note. Other stimuli included sets of visual stimuli following a logical pattern and the seven notes of an ascending scale in which the seventh note was either the "leading note" (seventh degree of a scale) or another note from the scale. Only the sentences resulted in a significant N400 effect. A later study by Paller et al. (1992) using similar stimuli in which a longer time also failed to find an N400 in response to musically incongruous elements, rejecting the idea that the absence of an N400 in the the study by Besson and Macar (1987) was due to the time course of the presentation of stimuli. Late negativities resulting from rule, memory and a combination of rule and memory violations were however reported in a music incongruity processing study by Miranda and Ullman (2007) though no mention was made of the delay between the stimulus and the response.

The first study to report an N400 in music was one by Hantz et al. (1997) which demonstrated an N400 component following the processing of the last note of a melody in which the musical phrase was not closed (ending on the tonic) but instead ended with a diatonic or chromatic open note. A burst of white noise presented instead of a note at that time led to a P300 effect, reinforcing the fact that an unexpected note within the musical system is processed differently to another unexpected element which results in an oddball effect — an ERP response to a stimulus which deviates from its context (e.g. a high tone embedded in a series of low tones, a capital letter in a series of lower case letters). However, close examination of the data revealed that the peak of the component occurred 500ms after the onset of that note, making the reported N400 seem more like the N500 reported by Koelsch, Maess, Grossmann, and Friederici (2003) when participants with no musical training processed an irregular chord at the end of a five-chord sequence. This late negative component, interpreted by Koelsch, Maess, et al. (2003) as an indicator of harmonic integration processes was not found when participants listened concurrently to sentences ending with an unexpected word and the chord sequences described above, when participants were asked to focus on the semantics of the stimuli (Koelsch et al., 2005). This could have been interpreted as a masking of the N500 by a concurrent N400 elicited by the processing of semantically unexpected words. However, the authors found that when semantically unexpected words were presented concurrently with a physically deviant musical tone, a significant N500 was generated. These results were taken to mean that the lack of an N500 when harmonic and semantic incongruities were presented concurrently could be attributed to the fact that the integration of the deviant chord was impeded by the integration of semantic information, which was easier and as such became the dominant task (Koelsch et al., 2005). This effect of concurrent semantic incongruity processing on the amplitude of the N500 generated by incongruous chords has been replicated in a study which showed that the concurrent processing of syntactic incongruities had no such effect (Steinbeis & Koelsch, 2008b).

The first study to report an N400 effect in music which was statistically indiscernible from the N400 effect observed in language was carried out by Koelsch et al. (2004) using a priming paradigm. In this controversial study, ERP effects were measured in response to a word (e.g. *flea*) following a sentence (e.g. about kings and princes) and following a musical extract (e.g. a Beethoven symphony). However, the N400 reported was in response neither to musical elements nor to a rule or probability violation within a musical stimulus but to a word presented after the stimulus. More recently, Steinbeis and Koelsch (2008b) demonstrated an N400 in musically trained participants on the processing of musical chords which were primed by words. Although the authors claim that their study is the first to show that a musical signal can elicit a meaning related N400, the meaning reported in these studies is arguably dependent on verbal associations within the listener's culture (i.e. the use of church music as a prime for the word "church" in the study by Koelsch et al., 2004; a pleasant chord as a target for the prime "love" and a dissonant chord following the word "hate" in the study by Steinbeis and Koelsch (2008a)). Furthermore, this N400 cannot be reported as a component which reflects the processing of a musical element within the symbolic system of music; instead, it seems to reflect a mismatch between the cultural associations of a musical element and a word. These findings do not therefore present an intra-music incongruity-integration based N400.

Further studies investigating semantic priming sought to provide a symmetrical effect. The study by Koelsch et al. (2004) used words as targets, and either music or sentences as primes. Daltrozzo and Schön (2009a) used both music as primes for word targets, to replicate the findings from the study by Koelsch et al. (2004), and sentences as primes for musical targets. In a desire to see how much musical information was necessary for a semantic concept to be evoked, the authors reduced the length of the musical extract to one second. These musical extracts were selected by musicians prior to the study to fit with the verbal labels provided by the authors. Participants were asked to perform a relatedness judgement task, with a yes or no button press, to indicate whether or not they thought the words were related to the music. ERP data showed a significant N400 in both conditions (music priming words and words priming music). No behavioural priming effect was found, where a faster response to related words would have been expected.

To investigate whether the N400 was task-driven, the authors conducted a second study (Daltrozzo & Schön, 2009b). Participants in the first study (Daltrozzo & Schön, 2009a) were specifically instructed to judge the relatedness of words to music — a task which would encourage the generation of semantic labels when processing the short musical extracts. In the new study, using the same stimuli, participants were asked to perform a lexical decision task, basing the investigation of priming on the paradigm developed by Meyer and Schvaneveldt (1971). In this new study, they found an N400 effect, but it was much weaker than when participants had been asked to perform a relatedness task. The authors concluded that the N400 was task-modulated, conceding that at least part of the priming effects were induced by the relatedness judgements. This study also showed no significant behavioural priming effects, where studies using lexical decision tasks have consistently found significantly faster reaction times to related words than to unrelated words (Daltrozzo & Schön, 2009b).

Considering the rapidity of the onset of the ERP component (beginning approximately 200ms after stimulus onset), the authors speculated that the "conceptual priming" they observed could be due to the timbre of the musical extracts. This explanation seems more plausible than the notion that less than 200ms of music could convey such concepts as magic, madness, wounding, courage, happiness, ugliness, speed and light. Indeed, traditional theories of musical meaning and emotions focus on purely intra-musical sources of meaning, and build upon the tension, anticipation, and the notion of resolution created by manipulating the relationships between musical elements governed by the notions of key and meter in a piece. This study points to an effect semantic association of timbre with specific concepts, akin to the association of specific types of music (e.g. church music) with specific words (e.g. church) in the original study by Koelsch et al. (2004).

A faster reaction time had been found in a study by Poulin-Charronnat, Bigand, Madurall, and Peereman (2006) in response to linguistically related words sung at the end of a chorale-type phrase than to unrelated words. This difference was modulated by the harmonic structure underneath that sung word, with more stable harmonic structures leading to a stronger linguistic priming effect. Despite the authors referring to this study as a priming study, this experiment does not present music priming language, but merely the processing of musical structure interfering with the processing of semantics, akin to the studies discussed in the next subsection (2.4.3).

Although the N400 was the principal sought component in the initial music and language ERP studies, both Besson and Macar (1987) and Paller et al. (1992) reported an ERP component resembling the P300 discussed above in response to incongruous endings to scales and melodies (Besson & Macar, 1987) and an early negativity (N100, Besson and Macar, 1987; N120, Paller et al., 1992) in response to incongruous final elements in stimulus sequences be those sentences, visual sequences, or melodies. Since this pattern occurred both after rule violations and memory violations, it was concluded that it was an indicator of incongruity detection, regardless of the reason for the incongruous status of the event.

Two early negativities evidenced in ERP literature have been disentangled in more recent studies. The first, the Mismatch Negativity (MMN), is a fronto-central negative component typically elicited by a discernibly different stimulus in a repetitive sequence of stimuli (Naatanen, 1976), such as a high tone embedded in a series of low tones. This effect, which could be equated with an auditory oddball effect, has been reported in response to repetition violation in music (Koelsch et al., 1999). The second, is a lateralised negativity, Early Right-Anterior Negativity (ERAN) in the right hemisphere and (Early) Left-Anterior Negativity ((E)LAN) in the left hemisphere. This lateralised component was disentangled from the oddball-type negativity by Miranda and Ullman (2007) in a study which showed a double dissociation between responses to repetition-based incongruities and rule-based incongruities defined above. This ERP response to rule-based violations has been reported in studies using non musicians (Koelsch, Maess, Gunter, & Friederici, 2001; Maess et al., 2001; Koelsch & Friederici, 2003; Jentschke et al., 2005), musicians (Jentschke et al., 2005; Miranda & Ullman, 2007) and children with varying degrees of musical and linguistic abilities (Koelsch & Friederici, 2003; Jentschke et al., 2005). This pattern mirrors the Left Anterior Negativity (LAN) which occurs in response to violations of language syntax (Koelsch & Friederici, 2003). Similar to the increase in amplitude of language-based components in response to an incongruity which is increasingly distant from expectations, the amplitude of the ERAN increases as the harmonic context from which listeners are able to predict the rest of a musical sequence increases (Koelsch & Friederici, 2003). This observation suggests a role for predictions in this early negativity, based on the incremental build-up of a stable representation of the stimulus according to the regularities previously encountered by the listener, distinguishing it from a more simple novelty-detection component such as the MMN.

A study investigating individual differences in ERP responses between children affected by Specific Language Impairment (SLI) and those with no language impairments (Koelsch & Friederici, 2003) demonstrated that both the ERAN and the (E)LAN are dependent upon the ability to process structure in language. Despite their rule-based nature, these early negativities appear to be independent from training and age, but dependent on the ability to process structure. Training effects were seen however in a third component, namely the Right Antero-Temporal Negativity (RATN). This component, which has a longer latency (around 350ms) has been seen only in the ERP data of musicians (Patel et al., 1998; Koelsch & Mulder, 2002) suggesting that the initial universal detection of an incongruity is followed by different processing dependent upon the extent of stored knowledge concerning the rules governing musical syntax.

The limited number of EEG electrodes used in early studies and the shared time window between the MMN and the ERAN make the differentiation of these two effects difficult in the studies by Besson and Macar (1987) and Paller et al. (1992). The authors however reported that the negativity was followed by an early positivity, describing the overall component as an N100-P200, or N1P2. This ERP pattern has since been reported in a seminal study by Patel et al. (1998) in response to music but not to language stimuli. This second part of the component has been repeatedly associated with the processing of a chord from the wrong key embedded within a probability-abiding chord structure (Patel et al., 1998; Maess et al., 2001).

A key contribution of Patel to the joint study of music and language cognition was the first demonstration of a statistically indistinguishable P600 occurring in response to syntactically improbable structures in language, such as embedded relative clauses, and chords belonging to one key embedded in a musical piece in a different key (Patel et al., 1998). This component echoes the Late Positive Component (LPC) reported in response to incongruous elements in musical melody (Besson, Faita, & Requin, 1994; Regnault & Besson, 1998; Besson, 1998), harmony (Koelsch et al., 2005), and rhythm (Besson et al., 1994; Miranda & Ullman, 2007), where the amplitude of the pattern increases with both familiarity and expertise (Besson et al., 1994). Regnault and Besson (1998) found that melody incongruities resulted in a LPC while semantic violations resulted in an N400. The LPC was seen to be longer in latency than the N400, regardless of whether participants were asked to focus on the music or on the language of the stimuli (extracts from opera songs). The amplitude of these components seem therefore both unaffected by the direction of attention and by the dual or single nature of the task performed (Besson, 1998).

2.4.2.1 Summary of ERP responses

Tables 2.1 and 2.2 shown at the end of this chapter present a summary of the responses reported in the literature in terms of the name assigned to them, their positive or negative deflection, their latency, their peak, the type of stimuli they occur in response to, and their interpretation given by the authors. These tables demonstrate some of the discrepancies between researchers in this field in terms of the names given to similar effects, and the distinctions drawn between similar effects by different authors. Nonetheless, several key responses can be summarised as follows.

The early negativities reported by authors seem to fall into four different effects. The first is most commonly referred to as the ERAN, though the lack of a right hemisphere preponderance led some authors to refer to it only as an Early Negativity (Steinbeis & Koelsch, 2008b) or Early Anterio-Central Negativity (Miranda & Ullman, 2007). This negativity is reported in response to incongruous chords and out of key violations of melodies (which would violate the underlying harmonic structure of a piece). It has been said to start as early as 130ms and as late as 290ms after the onset of the incongruity and to last until as early as 210ms to as late as 340ms after the onset of the incongruity, and has its source in BA44 in the right hemisphere (Maess et al., 2001). It occurs earlier in musicians than in non musicians, and earlier in response to very unexpected elements compared to mildly unexpected elements (Steinbeis et al., 2006). Interpretations of the ERAN range from the detection of a musical rule violation (Miranda & Ullman, 2007) to early syntactic processing (Koelsch et al., 2005) via expectancy violation (Steinbeis et al., 2006). Three other early negativities are the MMN, the RATN and the LAN. The MMN, frontal in its distribution, is interpreted as a response to physically deviant sounds and an indicator of preattentive auditory change detection (Koelsch et al., 1999). It has a latency of 275ms to 325ms with a peak at 300ms according to Koelsch et al. (1999), and a latency of 90ms to 190ms with a peak at 150 according to Koelsch et al. (2005). The differences in latencies and in scalp distributions reported in the literature make the intepretation of the Early Negativity reported by Besson and Macar (1987) and Paller et al. (1992) difficult. The RATN reported by Patel et al. (1998) in response to out of key chords embedded in musical sequences is said to reflect music-specific working memory processes and has a slightly later latency than the ERAN (300ms - 400ms). The lan, which occurs in response to syntactic violations shows a frontal left distribution, and a latency of 300ms to 400ms or 450ms with a peak at 390ms (Koelsch et al., 2005), and is said to reflect the early stages of syntatic processing. Despite the later latency of this effect compared to the ERAN, it has been reported as the linguistic counterpart of both the ERAN (Koelsch et al., 2005) and the RATN (Patel et al., 1998).

Later negativities can be broadly categorised according to their scalp distribution. The N400, most often found in response to low-cloze probability words and words at odds with the meaning conveyed by a music extract, has a latency beginning at 300ms and ending between 400ms and 600ms, and is said to reflect semantic processing. Late negativities in response to incongruous chords have been called modulations of N400s, N500s, and Late Negativity, and typically present a more anterior distribution, although the early study by Hantz et al. (1997) showed a more parietal distribution, and is reported to have its source in BA 21/37 (Koelsch et al., 2004). These effects typically begin between 450ms and 500ms after the incongruity and end 550ms to 900ms after the incongruity, and have interpretations ranging from rule-based harmonic incongruity detection to harmonic integration processes. A study by Koelsch, Gunter, et al. (2002) demonstrated a similar effect but with a much later latency (800ms to 1200ms) in response to chords which had been transposed at the end of expressively played extracts from classical piano music, suggesting that the qualities of the stimulus may affect the latency of this effect.

Early positivities have been reported with two different distributions. Early fronto-central positivities began between 200ms and 270ms and ended between 350ms and 400ms. These were found in most often in response to out-of-key harmonic or melodic violations and violations of familiar tunes, and were interpreted as a reorientation of attention to novel stimuli, and a function of anticipatory processes. Though it was equated with a P300 effect, this effect has typically been parietal (Jentzch & Sommer, 2001), making it more in line with a second type of early positivity reported in these studies said to reflect the processing of physically deviant sounds such as white noise (Hantz et al., 1997) and timbre change (Koelsch et al., 2005).

Though Besson and Macar (1987) reported a P300 in response to incongruous notes in melodies, and interpreted this component as an indicator of harmonic processing, its distribution (centro parietal), long latency (300ms - 1400ms) and peak (at 700ms) showed a strong resemblence to the Late Positive Component reported in later studies. Indeed Besson et al. (1994) reported a centro-parietal ERP effect with a latency of 300ms to 1400ms in response to harmonically and melodically incongruous notes as a Late Positive Component (LPC). This same study also showed a LPC in response to rhythmic incongruities, but this effect had a shorter latency (300ms to 700ms). This effect, also called a Late Posterior Positivity, has been reported in response to in-key, out-key (Miranda & Ullman, 2007), rhythmic (Besson et al., 1994), melodic (Besson & Macar, 1987; Besson et al., 1994; Miranda & Ullman, 2007) and harmonic violations (Koelsch, Gunter, et al., 2002; Koelsch et al., 2005; Regnault & Besson, 1998), and is said to reflect attention processes (Miranda & Ullman, 2007), integration difficulty (Besson et al., 1994) or decisions relating to the dissonant nature of a chord (Regnault & Besson, 1998). It begins between 300ms and 500ms post incongruity and ends between 500ms and 1400ms post incongruity. This component shows striking similarities with the typical P600 resported in response to linguistic syntactic incongruities (Koelsch et al., 2005; Steinbeis & Koelsch, 2008b), which begins between 300ms and 500ms and ends between 700ms and 1100ms with a similar scalp distribution. The direct comparison of music-elicited P600-type responses with language-elicited P600-type responses was carried out by Patel et al. (1998) who showed that despite a difference in the time-window used in the analyses of responses to music and language, the late positivity was statistically indistinguishable between domains.

The differences between stimulus sets used by the authors reporting statistically indistinguishable neural responses to music and language could be to blame for the discrepancies in the results. An explanation linking the N400 to a lack of harmonic closure rather than to the integration of the incongruity itself is able to account for these discrepancies in findings. Equally, the N400 in semantics could be interpreted as an indication of a lack of semantic closure, since the sentence does not make sense as a whole. The types of stimuli used by Patel et al. do not prohibit the closure of harmonic phrases, despite a difficult integration of the incongruous chord, due to a complex relationship between that chord and the key of the piece in which it was placed.

However good the fit of such speculations may be to the data presented in this review, the reliance solely on neuroimaging data and on a posteriori considerations of stimuli is fraught with confounds. One potential confound, resulting in language-related ERP components as a result of listening to music, is the fact that instructions were the inevitable influence of verbal descriptions of the stimuli and of participants' tasks on semantic activation during music. Another confound is the untested nature of the assumptions necessary for the conclusions to be reached, with researchers claiming that the statistically indistinguishable nature of ERP effects across music and language "suggests that [these components reflect] the operation of mechanisms shared by both linguistic and musical processes" (Poulin-Charronnat et al., 2006, p. B68). Other assumptions which should be the object of further investigation include the degree to which theoretical incongruities are perceived as odd, the basis upon which participants are making oddity judgements (novelty detection vs. a lack of closure), and the extent to which similar patterns, both behavioural and neural, can be equated with similar processes. Researchers agree that there needs to be a more unified approach to the study of incongruities in music and language (Juslin & Västfjäll, 2008; Vuust & Kringelbach, 2010; Thierry et al., 2008) in order for solid empirically substantiated conclusions to be reached. The research presented in this thesis aimed to contribute to this area by providing a set of rigorously developed stimuli and behavioural measures, and by addressing some of the assumptions relied upon in this area of research.

2.4.3 Similar patterns, shared resources?

Neuroimaging studies are often called upon to support the notion of brain regions being shared across domains, with the assumption that these are using the same processes and the same resources to accomplish the demands of the task under experimental investigation. Studies have shown for instance similar activation patterns in the processing of deviant stimuli as compared to normal stimuli in both domains (Tervaniemi et al., 2006). However important the biological plausibility of theories of cognition, the mere observation of shared neural activation patterns cannot be used to deduce an identity of processes taking place across both domains. Computational and psychological support must also be sought for theories of shared processes and resources across music and language.

Observing that error sensitivity did not differ significantly between single and dual task phases of their music and language experiment, Bonnel et al. (2001) concluded that incongruities in music and language are processed independently, since the use of cognitive resources for the processing of one stimulus type did not affect the processing of the other stimulus type. However, when I ran an ANOVA on the reported mean percent correct trials for the conditions “music alone”, “lyrics alone” and “music and lyrics” it showed a significant effect of condition, with mean percent correct trial in the dual task being significantly lower than mean percent correct trial in the two other conditions ($p < 0.05$). Thus it appears that engaging in the task of detecting both melodic and semantic incongruities at once does diminish the ability to detect incongruities compared to listening to the same materials in a single task. Looking at the data in this way suggests to the contrary that these tasks share resources, even if these are merely at the level of attention or output. The pattern emerging from studies reported in the previous section however highlights the importance of structural integration and the way in which dual tasks which both require a level of structural integration tend not to be performed as well together as they are separately.

A series of studies conducted by Koelsch and his colleagues comparing the ERAN and the (E)LAN lends strong support to the notion of shared mechanisms. Though these patterns originate from different hemispheres, there is an interference of one upon the other in that the elicitation of an ERAN concurrently with an (E)LAN decreases the amplitude of both components compared to their non concurrent elicitation (Koelsch et al., 2005). The facts that both components are absent in sufferers of SLI (Jentschke et al., 2005) and that the presence of an ERAN is seen at an earlier age in children having received musical training than in children who have not received any musical training (Koelsch, Grossmann, et al., 2003; Jentschke & Koelsch, 2009), suggest that a central structure tracking mechanism subserving both music and language is unable to detect incongruities in these listeners.

A study by Besson (1998) lends support to this interpretation of the data. The study showed no interference of semantic processing indicated by the negativity component on the amplitude of a concurrent LPC. This negativity, described by the authors as an N400 in fact begins 50ms after the onset of the incongruous word and before the component described as a P300 which peaks 600ms after the onset of the same word. The early onset of this negativity and the following independent prolonged negativity (which resembles an N400) and positivitity (resembling a P600) support a model in which a central incongruity detector or structure tracker exists which relies on distributed information and resources.

The fact that carrying out two tasks simulataneously will always lead to different effects than when the tasks are carried out separately could be offered as a confound in these studies. However, a recent study by Slevc, Rosenberg, and Patel (2009) addressed this issue with a thorough experimental design using garden path sentences and chorale-type musical sequences. Garden path sentences contain syntactically ambiguous elements which make the processing of the sentence structure difficult. Typical examples are "The horse ran past the barn fell. The log floated past the bridge sank. The ship sailed round the Cape sank. The old man the boats." (Harley, 2001, p. 248). In this study, participants read garden path sentences in a self-paced reading task, where each section of the sentence, typically one or two-words long, was accompanied by a chord from a chorale-type sequence. This method allowed the pairing of specific words with specific chords, such as a word which presented a syntactic difficulty with a chord which destabilised the harmonic structure of the chord sequence. These chords were of the type used by Patel and his colleagues (not preventing the closure of the harmonic sequence) previously associated with a P600 effect (Patel et al., 1998; Patel, 2003). The authors also manipulated the degree of semantic expectancy by using some sentences in which a harmonic incongruity was combined with a low-cloze probability word. Finally, stimuli in which the musical incongruity was the timbre of the chord underlying the critical word were created (a pipe organ sound rather than a piano sound).

The results showed participants spent longer reading the syntactically difficult word when it was paired with a harmonic incongruity, suggesting that the processing of harmonic structure and linguistic syntax relies on shared resources. No such effect of harmonic incongruity was seen in the processing of semantics, showing the specificity of the effect to syntax. Authors have argued that more basic attentional processes could still be responsible for this effect, since attention has been shown to be captured to different degrees by different chords within a given key (Jones, 1987). This study also provided a means of testing this explanation, by pairing syntactically difficult words with attention-grabbing changes in timbre and with harmonic incongruities in a second experiment. No effect of timbre change was seen on the processing of garden path sentences, but the effect of harmonic structure on the time spent reading difficult words in the sentences was replicated. This finding strengthens the conclusion that the effect on sentence processing was due to the concurrent processing of harmonic incongruities.

We are once more led to the SSIRH (Patel, 2003), which seems as yet the most plausible explanation for the data patterns described above.

2.4.4 Interpreting ERP components

A P600 component has been evidenced in the processing of erroneous mathematical equations (Martin-Loeches et al., 2006). Thus, the processes it denotes, which seem to be the integration of an incongruous element into an ongoing rule-governed stimulus, cannot be restricted to music and language cognition. This lends support to the portrayal of certain areas of the brain, previously considered central only to language processing, as structure trackers and integrators (Levitin & Menon, 2005), especially considering evidence of activation of BA 47 in the processing of arithmetic (Martin-Loeches et al., 2006).

In a review of neuroimaging data anchored around the SSIRH, Patel (2008) used differences in the instructions given to participants to suggest that the P600 could be attention-dependent. Studies by Koelsch and his colleagues, which have typically not found a P600 but instead have found early and late negativities have usually involved participants listening to the stimuli in a more passive way. Another explanation for the differences between the initial study by Patel et al. (1998) and the components observed by Koelsch, Gunter, et al. (2003) lies in the differences between the stimuli used by the two research teams. In stimuli typically used by Patel and his colleagues, the unexpected chords embedded within an ongoing sequence present several concurrent violations: a) they belong to another key, and b) the progression to and from these chords violates the rules of tonal harmony by using consecutive fifths and octaves (meaning that two or more voices in the harmonisation of a melody move in a parallel motion which violates the rules of classical harmony). The unexpected chords used by Koelsch and his colleagues present different characteristics: a) they are chromatic chords, which contain notes borrowed from a different key but are perfectly acceptable chords within the key of the sequence, albeit less probable than non-chromatic chords, b) the chord progression to and from these chords is in line with the rules of tonal harmony, and, most importantly to this discussion, and c) they are presented as the final chord of a five-chord sequence, and as such leave the sequence in need of harmonic closure (e.g. the sequence ends on an interrupted cadence or a chromatic chord). These chords have been widely used by composers since the 18th century as a means of deferring closure by delaying the fulfillment of harmonic expectations (Lovelock, 1957).

These two sets of stimuli were therefore designed to investigate two different types of violations. In the first, the violation is a chord from the wrong key, be it close or far. In the second, participants' implicit expectancies are violated within the rules of tonal harmony. Furthermore, the first type of violation is conceived as an error in an ongoing closed sequence whereas the second type leaves the chord sequence unfinished. These two types of incongruities can thus be conceived as follows: a syntactic violation resulting in a P600 component and a violation of the probabilistically-expected content of a syntactically acceptable sequence leading to a lack of closure. This distinction is in line with the interpretation of the late negativity offered by Hantz et al. (1997)

and stands up to more theoretical scrutiny than Koelsch and his colleagues' claim of semantic processing elements in music. Nonetheless, the interaction of semantic processing with the amplitude of the ERAN and the N500, and the observation of no N500 in sufferers of SLI (Jentschke, Koelsch, Sallat, & Friederici, 2008) is problematic to a dismissal of Koelsch et al.'s claims of semantic processing in both domains.

The reconciliation of these two schools of thought can be achieved when one considers the effect of semantic anomalies upon the incremental construction of a narrative. A sentence such as "He butters his bread with his..." leaves the participant expecting the word "knife" to complete the picture and bring closure to the message. By swapping the word "knife" for the word "socks", the picture is not complete, and instead, a different path in the narrative is opened such as an explanation for this odd behaviour, or a revaluation of the mental picture built up so far. Thus if we construe the N400 as an indication of incompleteness (in the narrative or in the musical sequence), this interpretation can capture both sets of data, allowing for the observation of interaction of semantic processing (narrative closure) with the processing of open-ended chord sequences (harmonic closure). Furthermore, with a semantic violation, there is no violation of the syntactic frame into which semantic content is introduced. In the same way, there is no disruption of the feeling that the chord which is about to be heard is the final chord of the sequence, simply a replacement of that chord with one which does not fulfill the closure function which a chord in this position should fulfill. This accounts for the lack of interference between syntactic processing and harmonic processing reported by Koelsch (2008). In the case of the chords used by Patel and his colleagues, the main key of the piece is disrupted by the insertion of a chord which should not be used in that key, leading to a shift into a new frame rather than an integration into an ongoing frame. This notion of integration and closure disruption is key to theories of music and language aesthetics which are founded on a striving for stability.

2.4.5 Step 2 — summary

One principle criticism of the ERP literature lies in the limited validity of stimuli and methods used as evidence to support the bold claims of equivalence between processes taking place in the processing of music and language. Nonetheless, from the literature explored so far, the following pattern emerges: music and language, processed at the level which requires structural integration, show clear similarities. Regardless of the explanation for the N400/N500 interaction, the literature reviewed above suggests both similar patterns in response to the violations of hierarchical structure and the coherence of the elements within that structure, and a degree of shared resources in the processing of these structure and their violations in both music and language. This provides the conditions necessary for the progression to the next stage in this review, to assess the evidence of shared affective and aesthetic patterns in response to incongruities in music and language.

2.5 Step 3: Can the aesthetic-affective responses to music and language be compared?

The literature explored so far points to strong similarities in music and language cognition, and to common patterns in the detection and integration of incongruities across these two domains. These observations, combined with the strong similarities shown in theories of aesthetics across music and language, lend support to the plausibility of similar mechanisms underlying the role of incongruities in aesthetics across domains. Indeed, the mechanisms underlying the models described by Meyer (1956) and Ortony (1975) and subsequent proponents of their theoretical stance show common patterns, namely the processing of music and language, and the way in which the brain deals with incongruities in both domains. It is therefore plausible to suggest that incongruities might act in a similar way in both domains to elicit aesthetic affective responses. However, one last condition must be fulfilled, namely the existence of similar affective and aesthetic responses to music and language.

2.5.1 Studies investigating affective responses in both music and language

Three studies recorded mood before and after exposure to music alone, lyrics alone (spoken) or music and lyrics together (Coffman, Gfeller, & Eckert, 1995; Gfeller, Asmus, & Eckert, 1991; Stratton & Zalanowski, 1994) with the common conclusion that both music and lyrics influenced participants' moods. The effect of a concurrent presentation of music and lyrics differed between studies. Coffman et al. (1995) found that the addition of music to lyrics increased the depressive tone of the text whereas Gfeller et al. (1991) stated that adding music to lyrics decreased the potency of the text. However, the same authors found that adding text to music increased the potency of the music in affecting mood ratings. Stratton and Zalanowski (1994) found that music alone did not significantly affect mood whereas the addition of lyrics helped the listener categorise the song as depressing or easy listening, significantly affecting mood ratings.

The direction in which the combined effect changed the mood (increasing or decreasing mood ratings by adding one stimulus to the other) varied between two experiments (Gfeller et al., 1991; Stratton & Zalanowski, 1994), leading to a problematic synthesis of findings. However, all three studies used only a small set of stimuli: one stimulus for Coffman et al. (1995), namely a song with depressing lyrics; one stimulus per condition for Stratton and Zalanowski (1994), undermining the validity of the experimental manipulation; one sad poem, *barbed wire* by R. H. Sauter (Gardner, 1965), set to one piece of music per condition for Gfeller et al. (1991). Consequently, the effects of music and language found in these studies, however significant, could be attributed to fixed effects of the stimuli themselves rather than to the experimental manipulations. Furthermore, no account is given for the mechanisms by which music and language affect mood, though all three studies concur that an increase in stimulus complexity (a simple melody line to a harmonised musical extract) is accompanied by an increase in the intensity of reported mood (Gfeller et al., 1991; Stratton & Zalanowski, 1994; Coffman et al., 1995). The generalisability of these findings and their relevance to the research question are therefore compromised, firstly because of the small number of stimuli used in each study and secondly because the focus of these investigations was on mood rather than on the immediate affective change in response to a specific element.

Four studies recorded participants' ratings of the stimuli rather than their moods following the presentation of the stimuli (Schmithorst, 2005; Ali & Peynircioglu, 2006; Steinbeis et al., 2006; Vuust et al., 2006). However, only Ali and Peynircioglu (2006) tested participants' responses to both music and language. As in the studies above, the stimuli consisted of music only, lyrics only or music plus lyrics. Participants used rating scales to rate their affective responses from 0 (not) to 9 (very) on four dimensions: 1) happy, joyful, exciting, festive, 2) sad, depressing, melancholy, 3) calm, relaxing, peaceful and 4) angry, unsettling, disconcerting, stressful. The authors found that affective responses were stronger to music than to lyrics. When participants listened to lyrics set to music where the intended emotion conveyed by the lyrics either matched or mismatched the intended emotion conveyed by the music, the melody was seen to weigh heaviest in the decision of what emotion the music-lyrics pair conveyed as a whole. Though interesting for the study of which component of a song has the strongest effect on aesthetic affective responses, this study does not provide insight into the way in which these components affect emotions. Furthermore, the effect of music and language on eliciting emotions is conflated with the perception of the conveyed emotion. These studies were therefore insufficient for providing evidence towards the research question of this review.

2.5.2 Sparse evidence for incongruity-induced affective responses

In both music and language, a link between expectations and emotions has been suggested by many authors (Meyer, 1956; Ortony, 1975; Ullman, 2001; Steinbeis & Koelsch, 2008b; Coulson & Van Petten, 2002; Huron, 2006), though conclusions concerning its existence often remain of a speculative nature. Of the twelve studies found in music and language psychology literature which claim to study the processes underpinning emotions in both domains, only seven report data pertaining to participants' emotional state while the remaining five draw their conclusions from neuroimaging data. For this reason, an exception to the criterion of studies needing to explore both music and language was made in this section. Evidence from recent studies investigating the role of incongruities on aesthetic affective responses to music and language was sought by carrying out a new search. The terms used were (music or language) and (affect* or aesthetic* or emotion*) and (incongruit* or expectanc*) in the title, keywords or abstract. The search yielded 107 publications, of which six were relevant to music and five to language.

Of the six studies relevant to music found in this search, only four reported new empirical data. Koelsch, Kilches, Steinbeis, and Schelinski (2008) recorded ERPs, skin conductance response and heart rate data while participants listened to naturally occurring unexpected chords in piano music. Participants were to rate how unpleasant the stimulus was, how aroused they felt and how surprised they felt at the end of the stimulus. The analysis however focussed on the role of emotional musical expression in playing and its lack of influence on ERP effects (ERAN), rather than on the effect of incongruities themselves on participants' affective responses. Previous studies also emphasised the role of para-musical cues such as stress or emphasis (Lindstrom, 2006) and timbre (Balkwill & Forde Thompson, 1999) in the eliciting and conveying of emotions within and across cultures. Finally a study by Müller, Höfel, Brattico, and Jacobsen (2010) reported participants' judgements of the beauty and grammaticality of musical stimuli, in which the final chord of a five-chord sequence had been modified to form a congruous, incongruous or ambiguous cadence, using a yes/no button press for each judgement. However the analysis of the data focused on the differences between expert musicians and lay persons within each congruence condition rather than on the effect of congruence on judgements and ERP effects.

In the six language studies, the notion of incongruity resolution as the basis of humour was mentioned a number of times (Bartolo, Benuzzi, Nocetti, Beraldi, & Nichelli, 2006; Cunningham & Derks, 2005; Galinanes, 2005) but of these studies only one study provided empirical data from a verbal task. Cunningham and Derks (2005) presented participants with jokes of varying comprehension difficulty and noted an inversely proportional relationship between how funny a joke was judged to be and the latency with which the joke nature of a statement was detected. The two other studies focused on laterality (Kutas & Hillyard, 1982) and prosody (Paulmann, Pell, & Kotz, 2008).

It would seem therefore that the emotions evoked by music and language show a potential for overlapping qualities but that more empirical research is needed both within each domain and across domains.

2.5.3 On the role of incongruities and structure tracking in both music and language

To date, only one study has set out to establish a link between incongruities processed in language and music, neural imaging and affect (Steinbeis et al., 2006). Although no data were collected using language stimuli, the patterns observed with the music stimuli were interpreted in the light of strong theoretical advances in psycholinguistics. This study used music which manipulated extracts from Bach chorales containing slightly unusual chords (chromatic chords), to make the stimuli more (out of key) or less (no chromatic chord) in line with musical expectancies. Results showed a significant increase of overall tension reported by participants in line with an increased departure from the norm for the manipulated chord. Electrodermal activity (EDA) also showed a significant increase between expected and unexpected chords (but no significant increase between unexpected and very unexpected chords). EEG data showed a significant effect of chord type on the eliciting of ERAN-type responses and late positivities. The authors concluded that their data support a link between musical tension, emotion and incongruities, and speculate that the perception of musical tension (created by a departure from the expected pattern) leads to higher arousal which in turn increases overall emotionality ratings. This causal chain remains in want of further investigation.

Exploring the relationship between incongruities and affective responses to music, Schmithorst (2005) linked the higher intensity of emotion ratings in harmonised pieces to the greater degree of expectation created by harmony. Though rarely directly investigated (e.g. Steinbeis et al., 2006), the relationship between the thwarting of expectancies and emotional responses is supported by the combination of evidence from disparate studies. BA 47 has previously been associated with structure tracking (Levitin & Menon, 2005), a process key to establishing predictions and experiencing resolution of thwarted predictions. The importance of structure tracking in establishing musical meaning led Levitin and Menon (2005) to suggest a role for BA 47 in the affective processing of music. This area was later shown to be activated in the processing of bi-stable rhythmic stimuli when musicians were asked to keep track of the original beat in a syncopated part of a musical piece (Vuust et al., 2006). Verbal reports of increased tension during the part of the musical piece in which the original metric structure was difficult to follow were noted by the authors. Thus, rather than being involved in musical meaning per se, as suggested by Levitin and Menon (2005), BA 47 appears to be involved in the tracking of structure, and thus helps create the tension experienced when the established structure is distorted or difficult to follow due to a counter rhythm. None of the studies reported above venture an explanation as to the emotional potency of poetry and song lyrics. However, evidence for a thwarting of expectations in metaphorical language (Coulson & Van Petten, 2002, 2007) and in syntactic emphasis devices such as the functional shift (Thierry et al., 2008) suggests that the same process of structural violations may underlie the emotionality of certain aspects of language.

Although these studies investigate either music or language, the similarities in neural patterns observed in studies investigating the processing of both types of stimuli has led some to conclude that identical processes are taking place in the two domains (Brown et al., 2006; Ducreux et al., 2003). These purely imaging studies can be called upon to support the plausibility of such theories, but the combined investigation of neural, behavioural and affective responses to music and language remains in need of further work.

The question of whether greater activation in these structure-processing cortical areas could be an effect of heightened attention was addressed earlier and answered in the negative considering the data collected during the processing of highly complex unstructured sound (Levitin & Menon, 2003). These observations of heightened activation and its speculated association with emotions fits with the model developed in this review according to which affective responses are elicited when the structure tracker is made to work harder to provide closure in processing of ongoing music and language.

2.5.4 Step 3 — Summary

Although the evidence emanating from the independent investigation of music and language aesthetic affective responses elicited through the violation of expectations seems to suggest a similar mechanism across both domains, this theory has in effect not yet been the subject of rigorous empirical investigation. The studies with the strongest empirical grounding both found elements which suggest that one of the main contributors to tension in music is a structure-based mechanism which is also found in language (Steinbeis et al., 2006; Vuust et al., 2006). Though these data support a link between emotions and the thwarting of expectations, the applicability of Meyer's (1956) model of meaning and emotions to language is as yet untested, and remains in want of theoretically grounded empirical investigation. The review of the literature pertaining to this question cannot therefore draw conclusions as to the similarity of processes underlying affective aesthetic responses to music and language, though evidence from studies looking at music and language separately suggests strong similarities between domains.

2.6 Conclusions and research plan

The review of the literature presented above led to the following conclusions. Firstly, shared processes and even resources may underlie the perception, encoding and retrieval of music and language. This allows for the possibility that the way in which music and language stimuli are stored, and the way in which mental representations are accessed, are similar in both domains. This, in turn, allows a progression in the review to the next condition necessary for a cross-domain model of incongruity-based aesthetic-affective responses. Secondly, this review showed that expectations are demonstrated in both music and language, and the processing of elements which violate these expectations and learnt regularities results in similar neurophysiological and behavioural patterns. Authors argued for commonalities in the application of probability-based expectations to online music and language processing. Although the interpretation of certain ERP components is still under debate, the statistically indistinguishable neurophysiological patterns observed in the processing of both types of stimuli and the interaction between the two domains strongly suggests shared resources if not shared processes. Finally, the literature concerning emotions arising from music and language incongruities is sparse and for the most part includes only studies investigating music or language, not the two domains together. Nonetheless, the similarities found in the studies investigating aesthetic affective responses in music and language separately support incongruity-based affective responses in both domains.

Based upon these observations, the literature explored in this review gave rise to the following conclusion: yes, the same mechanism, underpinned by the detection and integration of incongruities, *could* underlie the eliciting of affective aesthetic responses in music and language. However, the scarcity of literature empirically investigating this possibility calls for further research in the area. It seems therefore that, although recent methodological advances permitted a more systematic investigation of incongruity-based models of music and language aesthetics, the empirical research necessary for the the veracity of a theoretically plausible shared mechanism to be assessed was still lack-ing. These observations provided the impetus for this thesis, in which the following research strategy was adopted not only to bring further understanding to the role of incongruities in aesthetic experiences of music and language, but also to help contribute to the unified approach yearned for by researchers studying incongruity processing and aesthetic affective responses both within and across these two domains of human cognition.

2.6.1 Can stimuli presenting equivalent levels of incongruity be created to investigate the role of incongruities in aesthetics in both music and language?

In view of the discrepancies mentioned between the stimuli used by different research teams in this field of research (Patel et al., 1998; Koelsch, Gunter, et al., 2003), and the calling into question of the validity of stimuli used to demonstrate the conveying of meaning by music (e.g. Koelsch et al., 2004), a first step in this research was to provide a theoretically-grounded and empirically validated stimulus set for the study of incongruities across domains. By ensuring a manipulation of both incongruity and resolution, the set of stimuli described in Chapter 3 allowed the investigation of whether the harmonic closure account given by Hantz et al. (1997) for the occurrence of late negativities in response to harmonic incongruities could explain the differences between responses reported in Tables 2.1 and 2.2.

2.6.2 Are theoretically odd stimuli perceived as odd?

The studies reported in Section 2.4.2 used theoretical distinctions to create "odd" musical stimuli, relying on cadence interruptions, rhythmic delays, chromaticism and on the circle of fifths to determine how unexpected a musical element would seem to the listener. Where in language studies degrees of incongruity are determined through completion studies and grammaticality judgements (Van Berkum et al., 2005), no such studies had been carried out in music, leaving unaddressed the question of whether theoretically odd elements are perceived as odd.

Experiment 1 (described in Chapter 4) presented participants with the stimuli designed for this project, to assess whether the experimentally manipulated degree of incongruity (*congruence*) in the stimuli resulted in differences in ratings of how odd the stimuli sounded. This experiment also served to refine the *congruence* condition fitness of the stimuli, in order to provide both theoretically-based and empirically-refined stimuli for subsequent experiments in this research, aligning with the requirements expressed in the first research question.

2.6.3 Can music and language elicit the same verbal responses?

The scarcity of research investigating music and language aesthetics in parallel described in Section 2.5.2 called for an investigation of whether similar responses could be found in both domains when a carefully controlled set of stimuli showing equivalence across music and language was used. The evidence reviewed suggested similar patterns could be found in both domains, in view of the similarity of the responses recorded in studies investigating music and language aesthetics separately from each other (Coulson & Van Petten, 2007; Huron, 2006; Steinbeis et al., 2006).

In Experiment 2 (described in Chapter 5), participants listened to the refined set of stimuli and were asked to type a word or two after each stimulus was played to describe their "gut reaction" to the stimulus. These verbal descriptions of participants' responses showed a certain overlap between music and language, allowing the creation of rating scales which could be used in both domains to capture different facets of participants' responses which were affected by both the music and language stimuli to which they were exposed.

2.6.4 How does the degree of incongruity affect participants' responses to the stimulus as a whole and to specific musical or linguistic elements of interest?

Despite the methodological advances allowing the non-invasive investigation of neural mechanisms associated with the processing of incongruities, and the possibility of recording many different types of data simultaneously, the literature presented in this review showed a lack of systematic parallel investigation of physiological effects (fMRI, ERP, EDA) and behavioural data (self-report, overt incongruity detection, ratings of emotionality). This left the field unable to address questions of the relationship between incongruities, the physiological effects linked to their processing, and their relationship to aesthetic and affective responses speculated upon in the literature.

Experiment 3 presented participants with the refined stimulus set while EEG data were recorded, time-locking ERPs to specific elements of interest in each stimulus, namely the onset of the incongruous element and its resolution. This method allowed the investigation of what neuro-physiological changes were associated with specific musical and linguistic elements, following from Meyer's (2000) plea for the advancement of his model.

The scales developed through Experiment 2 were also used in Experiment 3 to record participants' responses to the stimuli, and provided a means of assessing which element of participants' reported responses corresponded most closely with variations in EEG patterns. These also enabled the study of the effect of incongruity on different aspects of participants' responses, and the relationship between these different aspects.

2.6.5 Further questions addressed alongside the main research strategy

Several other questions arising from this review of the literature were addressed in this research. The question of whether the different responses described in Section 2.4

could be accounted for by the notion of closure (Hantz et al., 1997) was investigated by creating stimuli in which the degree of resolution as well as the degree of incongruity were manipulated. The data reported in Chapter 8 lends support to this notion. The divergence of interpretations of ERP components illustrated in tables 2.1 and 2.2 was addressed by collecting behavioural data on a variety of scales from oddity detection to confusion and tension alongside the collection of ERP data in Experiment 3. In view of the theoretical appeal of the SSIRH, the evidence of shared resources in music and language (Patel, 2003; Steinbeis & Koelsch, 2008b) and evidence showing effects of musical training on different aspects of human cognition (Forgeard, Winner, Norton, & Schlaug, 2008; Overy, 2003; Schlaug, Norton, Overy, & Winner, 2005; Tallal & Gaab, 2006), musicians and non musicians were compared on all aspects of this research to investigate any effects of musical training on music processing, and any effects transferred from musical training onto language processing.

2.6.6 Conclusion

This review provided a solid basis of plausibility upon which to build a cross-domain model of music and language aesthetics centering on the role of incongruities in eliciting aesthetic and affective responses. It also helped identify areas of research which required further investigation, and ways in which these could be investigated alongside the main research question of this thesis. Thus, the methods, theoretical developments, and resources used in this thesis were put to advantage, by building upon the investigation of a specific part of the overlap between music and language psychology to contribute more broadly to psychological enquiry.

Table 2.1: Summary of reported ERP effects

Name given	Reported by	In response to	Sample	Deflection	Latency	Peak	Topography	Said to reflect
ERAN	Koelsch et al. (2003)	Irregular sequence-final chord	Non musicians	Negative	130 - 280		Frontal	-
Early anterio-central negativity	Miranda and Ullman (2007)	Out of key violation of unfamiliar melody	Non musicians and musicians	Negative	150 - 270		Anterior central right	Detection of musical rule violation
ERAN	Koelsch et al. (2005)	Irregular sequence-final chord, eliminated by concurrent semantic incongruity	Non musicians	Negative	150-250	190	Frontal, slightly right	Early stages of musical syntax processing
ERAN	Steinbeis and Koelsch (2008)	Sequence-final incongruous chord	Non musicians	Negative	160 - 260	210	Right anterior	May reflect syntactic processing
mERAN	Maess et al. (2001)	Incongruous chord	Non musicians	Negative	170 - 210	200	Right Anterior	Harmonic context violation
EN	Steinbeis, Koelsch and Sloboda (2006)	Improbable chord embedded in piece	Musicians	Negative	180 - 230	200	Broad, Fronto central	ERAN, processing of harmonic expectancy violation
Negativity	Koelsch and Mulder (2002)	Transposed phrase-final chord in real music	Non musicians	Negative	180 - 350	250	Right, centro temporal	Harmonic incongruity detection, shown here in real music as well as artificially produced music
EN	Steinbeis, Koelsch and Sloboda (2006)	Slightly improbable chord embedded in piece	Musicians	Negative	190 - 240	210	Broadly distributed, Fronto central	ERAN
EN	Steinbeis, Koelsch and Sloboda (2006)	Improbable chord embedded in piece	Non musicians	Negative	200 - 250	230	Broadly distributed,	ERAN
Early posterior negativity	Miranda and Ullman (2007)	In-key and out of keyviolations in familiar melodies	Non musicians and musicians	Negative	220 - 380		Posterior	Detection of violation of memorised representation; N400, possible contamination through lyrics processing
ERAN	Jentschke et al. (2005)	Incongruous final chord	Children without SLI	Negative	230 - 350		Frontal, slightly right	Rule-based harmonic incongruity detection
Early negativity	Besson and Macar (1987)	Incongruous note in scale or familiar melody	-	Negative	50 - 150	100	Central, Parietal	Incongruity detection
N120	Paller et al. (1992)	Incongruous ending of familiar tune	-	Negative	80 - 160	120	Fronto-central	Allocation of additional attention
EN	Steinbeis, Koelsch and Sloboda (2006)	Slightly improbable chord embedded in piece	Non musicians	Negative	290 - 340	310	Slightly left anterior	ERAN
RATN	Patel et al. (1998)	Chord from different key to piece	Musicians	Negative	300 - 400		Right Anterio-Temporal	Similar to LAN, music- specific working memory resources
MMN	Koelsch et al. (1999)	Physically diviant tone in series	Non musicians	Negative	275 - 325	300	Frontal	Preattentive auditory change detection processes
MMN	Koelsch et al. (1999)	Mistuned chord embedded in series	Musicians	Negative	275 - 325	300	Frontal	Preattentive auditory change detection processes
MMN	Koelsch et al. (2005)	Physically diviant tone in series	Non musicians	Negative	90 - 190	150	Frontal	Early processing of physical deviants
LAN	Koelsch et al. (2005)	Syntactic violation (gender disagreement)	Non musicians	Negative	300 - 450	390	Frontal, left	Early stages of linguistic syntax processing
LAN	Steinbeis and Koelsch (2008)	Syntactic incongruity (gender violation) and low cloze words	Non musicians	Negative	300 - 400	-	Left anterior	-
N400	Daltrozzo and Schön (2009 a and b)	Short music extracts and words (relatedness task)	Non musicians	Negative	250 - 650	450	Parietal	Semantic priming (timbre associations?)
N400	Steinbeis and Koelsch (2008)	Low cloze probability word	Non musicians	Negative	300 - 400	-	Globally distributed	-
N400	Koelsch et al. (2005)	Low cloze probability word	Non musicians	Negative	300 - 450	350	Centro-parietal	Semantic processing
N400	Koelsch et al. (2004)	Words at odd with the meaning conveyed by music	Non musicians	Negative	300 - 500	410	Centro-parietal	Semantic priming by music
N400	Kutas and Hilliard (1980)	Low cloze probability word	-	Negative	300 - 600	400	Central	Semantic processing
N400	Daltrozzo and Schön (2009 a and b)	Short music extracts and words	Non musicians	Negative	500 - 650	550	Right, Frontal	Semantic priming

Table 2.2: Summary of reported ERP effects (continued)

Name given	Reported by	In response to	Sample	Deflection	Latency	Peak	Topography	Said to reflect
Late negativity	Hantz et al. (1997)	Musically "unclosed" melody	Musicians	Negative	-	471	Parietal	Processing of deviation within set of symbols
N500	Koelsch et al. (2003)	Irregular sequence-final chord	Non musicians	Negative	450 - 600	500	Frontal	Harmonic integration processes
N5	Steinbeis and Koelsch (2008)	Sequence-final incongruous chord	Non musicians	Negative	450 - 900	650	Right anterior	-
N5	Jentschke et al. (2005)	Incongruous final chord	Children without SLI	Negative	500 - 700		Frontal	Rule-based harmonic incongruity detection
Later anterior negativity	Miranda and Ullman (2007)	Out of key and in-key violation of melody	Non musicians and musicians	Negative	500 - 700		Anterior	Consistent with N5, rule and memory expectancy violation
Late negativity	Koelsch and Mulder (2002)	Transposed final chord (real music)	Non musicians	Negative	800 - 1200	-	Frontal, slightly right	Similar to N5, reflecting harmonic integration
P340	Paller et al. (1992)	Incongruous ending of familiar tune	-	Positive	200 - 400	350	Frontal	Equated with P300, processing low probability element
P3	Regnault, Bigand and Besson (2001)	Unexpected chord (I - IV cadence)	Musicians and non musicians	Positive	200 - 400	-	Fronto (central)	Anticipatory processes
Anterior-central positivity	Miranda and Ullman (2007)	Out of key violations in familiar melodies	Musicians	Positive	270 - 350		Anterior central	Consistent with P3, automatic orientation of attention to novel stimuli
Anterior-central positivity	Miranda and Ullman (2007)	Out of key violations in unfamiliar melodies	Non musicians and musicians	Positive	270 - 350		Anterior central	Consistent with P3, automatic orientation of attention to novel stimuli
Anterior-central positivity	Miranda and Ullman (2007)	In-key violations in familiar melodies	Musicians, larger than non-musicians	Positive	270 - 350		Anterior central	Consistent with P3, automatic orientation of attention to novel stimuli
P300	Hantz et al. (1997)	White noise at end of melody	Musicians	Positive	-	273	Parietal	Processing of physically divergent stimulus
P300	Koelsch et al. (2005)	Physically diviant tone in series	Non musicians	Positive	300 - 400	345	Right, parietal	Processing of physical deviants
P300	Besson (1998)	Incongruous note in melody	Musicians	Positive	300 - 1400	700	Central, Parietal	Harmonic processing
Early Positivity	Steinbeis, Koelsch and Sloboda (2006)	Improbable chord embedded in piece	Both	Positive	300 - 400	340	Central	Attention shifting
P600	Koelsch et al. (2005)	Syntactic violation (gender agreement)	Non musicians	Positive	450 - 700	500	Centro-parietal	Syntactic integration
P600	Steinbeis and Koelsch (2008)	Syntactic incongruity (gender violation) and low cloze words	Non musicians	Positive	500 - 600	-	Globally distributed	-
P600	Patel et al. (1998)	Grammatical difficulty	Musicians	Positive	300 - 1100	850	Posterior, midline	Increased processing cost
P600	Patel et al. (1998)	Chord from different key to piece	Musicians	Positive	500 - 800	600	Posterior, midline	Integration difficulty
Late Positive Component	Besson et al. (1994)	Harmonic or melodic incongruity	Musicians and non musicians	Positive	300 - 1400	-	Central, Parietal	Integration difficulty
Late Positive Component	Besson et al. (1994)	Rhythmic incongruity	Musicians and non musicians	Positive	300 - 700	-	-	Integration difficulty
Late Positive Component	Regnault and Besson (1998)	Unexpected chord (I - IV cadence)	Musicians and non musicians	Positive	300 - 800	-	Globally distributed	Effect of harmonic context, classification as dissonant
Late Positive Component	Koelsch and Mulder (2002)	Transposed final chord (real music)	Non musicians	Positive	350 - 1200	500	Posterior	Structural integration
Late Positivity	Besson and Macar (1987)	Incongruous note in scale or familiar melody	-	Positive	350-650	400	Parietal	Harmonic processing
Later Positivity	Steinbeis, Koelsch and Sloboda (2006)	Improbable chord embedded in piece	Musicians	Positive	400 - 500	470	Posterior	Similar to P300, incongruity detection
Late Positive Component	Koelsch et al. (2005)	Irregular chord concurrent to word	Non musicians	Positive	450 - 700	500	Centro-parietal	-
Late posterior positivity	Miranda and Ullman (2007)	In-key violation of familiar melody	Musicians	Positive	500 - 700		Posterior	Attention allocated to violation
Late posterior positivity	Miranda and Ullman (2007)	Out of key violation of unfamiliar melody	Non musicians and musicians	Positive	500 - 700		Posterior	Attention allocated to violation
Late posterior positivity	Miranda and Ullman (2007)	Out of key violation of familiar melody	Musicians	Positive	500 - 700		Central and Posterior	Attention allocated to violation

Chapter 3

Systematic oddities: the creation of stimuli to investigate incongruity processing in music and language

3.1 Introduction

In Chapter 2, concerns were raised over the validity of the stimuli used in the study of music and language cognition, which could account for the lack of advancement in this area of research. Firstly, since seminal music and language ERP studies were not designed with aesthetics in mind, stimulus sets were developed around rule-violating incongruities (Patel et al., 1998; Koelsch et al., 2001) rather than aesthetically-purposed incongruities (although see Steinbeis et al., 2006, and Koelsch, Kilches et al., 2008, for examples of stimuli derived from J. S. Bach's aesthetic incongruities). Secondly, studies which collected affective aesthetic responses were conducted separately for music (Steinbeis et al., 2006; Vuust et al., 2006) and language (Coulson & Van Petten, 2007; Thierry et al., 2008), leaving theories of shared processes at a speculative level. Thirdly, the stimuli used in music ERP research have been seen to violate the rules of classical harmony. The stimulus sets used by Patel and collagues for instance (see examples in Patel et al., 1998 and Slevc et al., 2009) contain consecutive fifths and octaves, while the placement of chromatic chords as a substitution for the tonic after a dominant by Koelsch and colleagues rather than a substitution for the dominant itself lead to erroneous cadences (see examples in Maess et al., 2001). These stimuli confound the study of experimentally controlled incongruities with other aspects of musical grammar. Fourthly, the aesthetic models described above rely on the notion of resolution. However, stimuli have typically consisted of sentences and musical sequences in which a specific element is changed across conditions to make it normal, improbable or erroneous. This leaves the resolution of the incongruity unaddressed, and does not enable the distinction between whether the heightened aesthetic experience is a result of novelty, or, as suggested by music theory, a result of a pleasant resolution and integration of a surprising element into the incrementally-built musical structure. A more suitable approach would be to use stimuli in which the degree of resolution of the incongruous event varies. Finally, where psycholinguists would use norming studies and corpus evidence to judge the degree of incongruity of their stimuli (Van Berkum et al., 2005), authors investigating music incongruities have tended to rely solely on a theoretical departure from the norm, leaving unaddressed the question of whether theoretically odd elements are consciously perceived as odd.

This chapter describes the development of a stimulus set in which the level of incongruity (or *congruence*) was systematically manipulated in an equivalent way across music and language, starting from incongruities used by authors and composers for aesthetic effect. To do so, in the first *congruence* condition, the context of the original aesthetically-purposed incongruity was either kept as such, leading to "incongrous-resolved" stimuli which contained an unusual element which resolves into the preceding context of the sentence or musical piece, typical of the stimuli discussed in the original incongruity-based theories of music and language aesthetics. In a second *congruence* condition, the context of the incongruity was changed to accommodate the incongruity so that it is no longer incongruous to its context, creating stimuli which are "normal" despite containing the same element of experimental interest as the first *congruence* condition. In a final *congruence* condition, the beginning of the stimulus was kept as in the original, but the end was or changed to fit with the incongruity creating a mismatch between the beginning and the end of the stimulus — the "incongruous-unresolved" condition.

During this creation process, care was taken to address the first four concerns expressed above. The first concern to do with the aesthetic nature of incongruities was addressed by starting from aesthetically-purposed incongruities used by authors and composers. The second concern, surrounding the equivalence of stimuli across domains, was handled by developing an equivalent manipulation across domains. The third concern, to do with the lack of good form of the stimuli was tackled by adhering to the rules of classical harmony and good linguistic form. Finally, the fourth concern, which emphasised the lack of suitability of previously used stimuli due to the lack of experimental variation of the resolution of incongruities, was addressed by introducing a manipulation of whether the incongruity resolved or not. This manipulation allowed the testing of whether the heightened aesthetic experiences reported in previous experiments were due to the novelty of a departure from the norm, or to the satisfaction of making sense of a stimulus as a whole in which a departure resolved in a meaningful way. The fifth concern is dealt with in Chapter 5 which describes a norming procedure based on participants' responses to an initial rating study (described in Chapter 4), and Chapter 7 which demonstrates the reliability and validity of the stimulus set.

3.1.1 A priori considerations

A significant amount of time was devoted to the creation of stimuli fit for this research project. To inform this process, both theoretical and empirical publications exploring parallels between music and language were reviewed. A table detailing these parallels was created, organizing the literature into specific types of music-language equivalence claims, such as notation, perception, hierarchical structure, parsing of sound, and incongruity processing (see Appendix A). Two key characteristics of music and language seen in this literature review are their combinatorial and grammar-abiding properties. The first characteristic means that in both domains a whole (e.g. musical phrase, sentence) is made up of a number of units (e.g., notes, words) which can be combined in a number of different fashions within a hierarchical system (e.g., notes, phrases, pieces; words, phrases, sentences). The second means that the arrangement of these units cannot be totally random but must abide by certain rules in order to create meaningful wholes (e.g. pieces with a consistent beat or key, sentences which have a sound syntactic structure). Thus, the overall structure of a musical extract can be equated to the structure of a language extract, with sequential small units hierarchically organized into larger units within which each element plays a functional role. This observation prompted the decision to focus on incongruities within the major structural underpinnings of music and language, namely syntax, semantics, harmony and rhythm.

A second table was used to establish whether any consensus had been reached concerning which units of the combinatorial music and language systems are directly equated with each other in the literature. Table 3.1 shows papers jointly investigating music and language cognition at the intersection of a row and column which show the units equated with each other across the two domains. For example, Besson (1999) reported similar effects for words in language and notes in music. The study is therefore placed in the table at the intersection of the *word* column and the *note* row. From this table, it appears that elements of harmony, from single notes within a key to intervals and chords have been associated with both semantic and syntactic elements in language, from words to discourse (Longuet-Higgins, 1994; Besson & Macar, 1987; Lim, 2006; Patel et al., 1998; Besson & Schon, 2001) and that a number of studies have reported parallels between the processing of chords or notes and single words (Besson, 1999; Patel et al., 1998; Besson & Schon, 2001; Koelsch et al., 2005; Jentschke et al., 2005). In view of these observations, the decision was made to focus on incongruities at the level of a word or a chord within the present research and to construct stimuli in which these incongruities were of a semantic or syntactic nature in language and a har-

Table 3.1: Equivalences between music and language in both theoretical and empirical publications

	phonemes	sequential events	phrase structure	words	semantics	syntax (gender /NV agreement)	nouns, verbs	subject / predicate	sentence	grammatical category	inflection	prosody	temporal violation	phrase (prosodic boundary)	no equivalent
pitches															
rhythm	Besson (2001)		Longuet-Higgins (1994)									Lim (2006)			
tonality															
note				Besson (1999)											
tone				Binder (1994) Besson (1987) Goodglass (1977)	Miranda (2007)										
harmony					Besson (1987)	Lim (2006) Patel (1998) Besson (2001)									
chords				Besson (1999) Lim (2006)		Koelsch (2005) Koelsch (2005b) Jentschke (2005)	Patel (1998)								
intervals				Lim (2006)			Patel (1998)								
tonic / dominant								Patel (1998)							
extract									Koelsch (2004)						
timbre											Lim (2006)				
temporal violation													Weyert (2001)		
phrase														Neuhaus (2006)	
octave equivalence															Patel (1998)
no equivalent										Patel (2003)					

monic or rhythmic nature in music. More specifically, the semantic stimuli developed for this research contained an incongruous word and the harmonic stimuli contained an incongruous chord. Rhythmic stimuli presented a change within the rhythmic structure, the onset of which could be clearly identified within a beat, and syntactic stimuli contained a single word which appeared to disrupt the sentence structure while remaining meaningful.

Although rhythm was not mentioned explicitly in connection with either semantics or syntax in the studies reviewed above, memory studies and neuroimaging evidence both suggest that the neural structures underpinning the processing of structure in language are also activated in the processing of rhythm in music (Levitin & Menon, 2005; Saito & Ishio, 1998; Saito, 2001; Vuust et al., 2006). Harmonic structure in music has been equated with both language syntax (Patel et al., 1998) and semantics (Koelsch, 2008). I make no claims that syntax is equivalent to harmony and semantics to rhythm or vice versa, at least at this stage in the investigation. My aim was to create cognitively equivalent incongruities across all four types of stimuli by applying the same experimental manipulation, described in the first section of this chapter, to the structural underpinnings of music and language, namely harmony, rhythm, semantics and syntax. The four-level variable denoting the type of stimulus is referred to as *hrss*; its four levels are *harmony*, *rhythm*, *semantics* and *syntax*. This approach enables researchers using these stimuli to establish whether systematic disruptions to the structural underpinnings of music and language have similar consequences in terms of processing costs, ERP patterns and aesthetic-affective responses.

3.2 Creating the stimuli

Before delving into the specifics of the stimuli, a few terms which will be used throughout the description of the creation and selection processes must be defined. These are illustrated in Figure 3.1 and described below:

hrss harmony, rhythm, semantics or syntax (h, r, se, sy): refers to the system which is disrupted by the incongruous element.

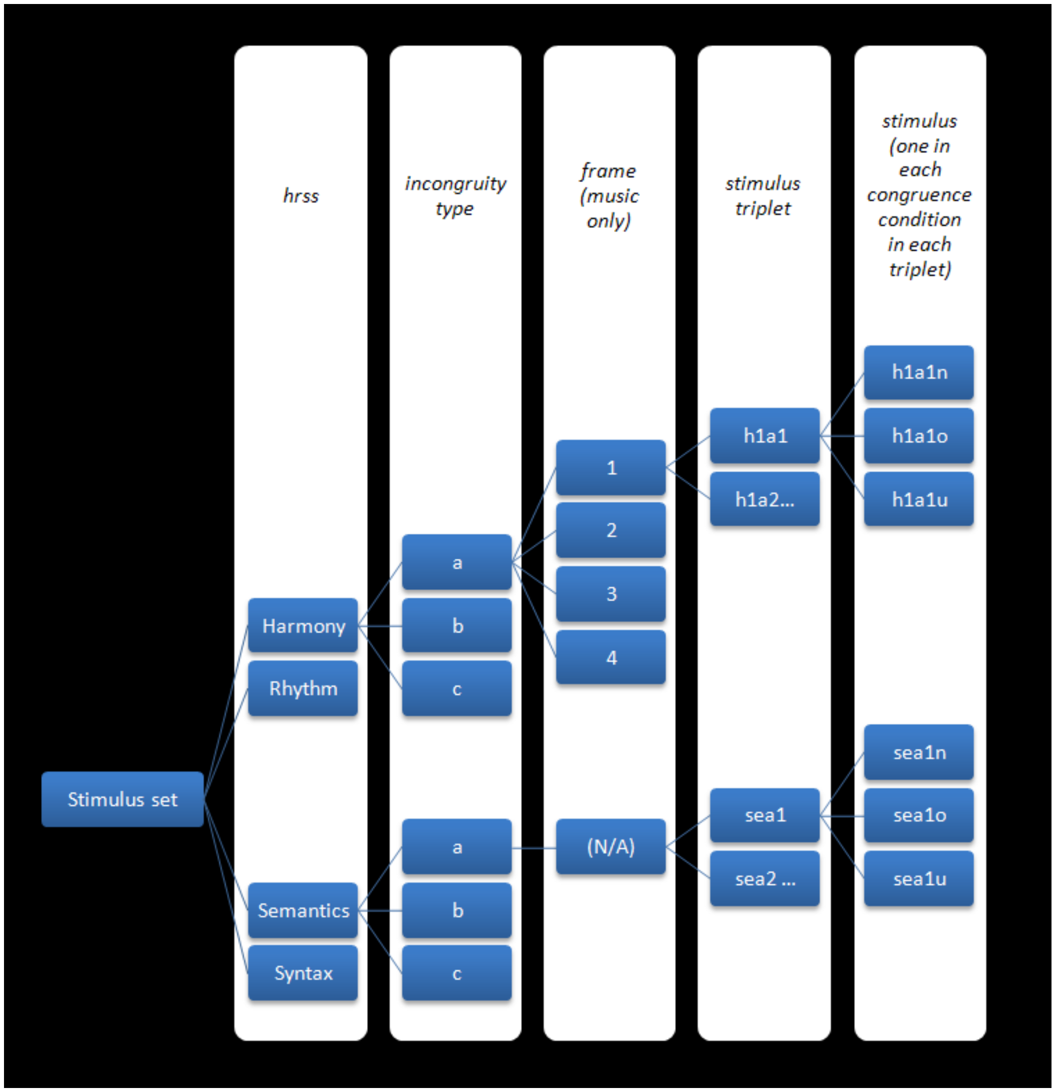

Figure 3.1: Definitions and illustration of key terms

congruence normal, incongruous-resolved, incongruous-unresolved: refers to the experimental manipulation carried out on a music or language extract; the specifics of the conditions are defined separately for each level of *hrss* in the following sections of this chapter.

incongruity type a, b, or c: refers to the degree of departure from the norm in the incongruous-resolved condition, with *a* being the smallest departure and *c* the largest departure; these departures are defined separately for each stimulus type in the following sections.

frame this term is only used with respect to the music stimuli: the rhythmic pattern (for the *harmony* stimuli) and the chord structure (for the *rhythm* stimuli) upon which the original incongruity (respectively odd chord or odd rhythm) is reproduced and manipulated (see sections describing the music stimuli for clarification).

stimulus triplet e.g. h1a3, seb1, r3c1, sya1: refers to three stimuli (one in each *congruence* condition) all derived from the same music or language extract; the name is composed of [hrss], [frame] (for the music stimuli), and [triplet number].

stimulus e.g. h1a3n, h1a3o, h1a3u: the name of each stimulus is made of the stimulus set plus *n* for the normal *congruence* condition, *o* for the incongruous-resolved condition, standing for *original* as these stimuli were the basis of each stimulus set, and *u* for the incongruous-unresolved condition.

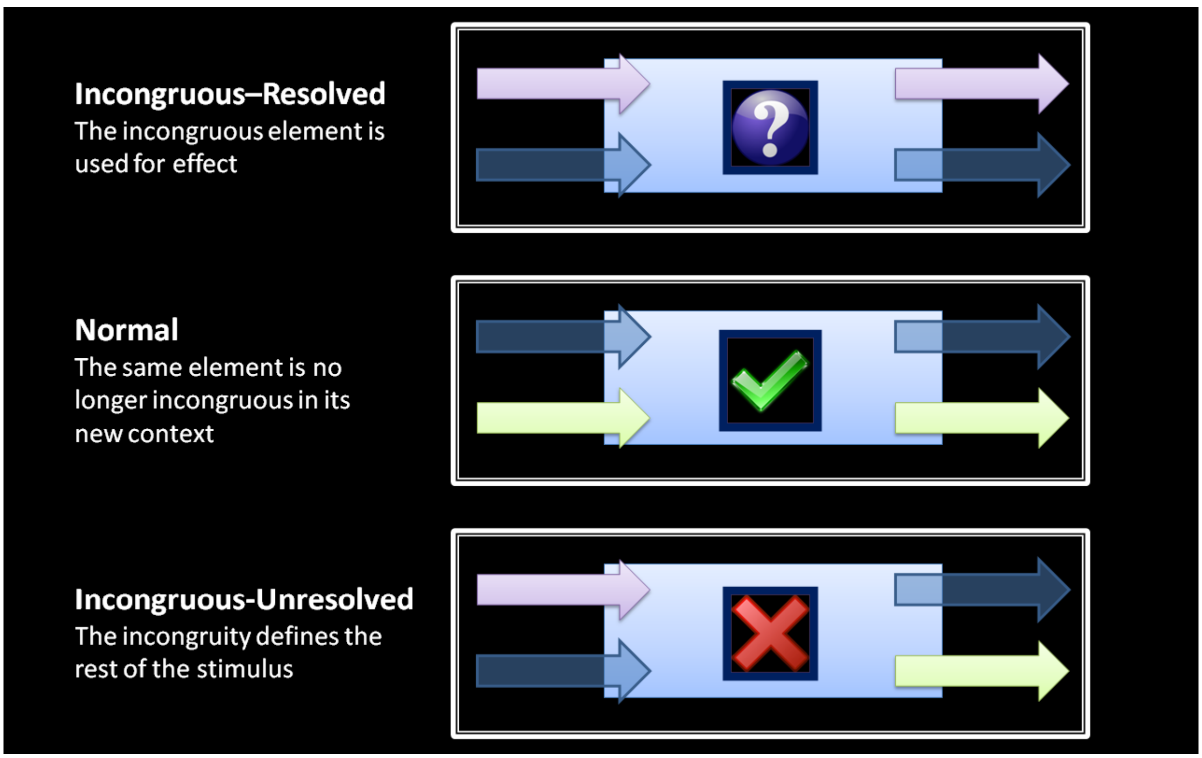

Figure 3.2: Format for the creation of all stimulus triplets in this research project

The *congruence* manipulation, illustrated in Figure 3.2 can be described as follows:

Stimuli described as *normal* do not present any elements (word, chord or beat) which are difficult to integrate into the ongoing structure of the stimulus (e.g. a chord sequence which uses only the most frequently used chords of a given key, or a sentence with a simple syntactic structure).

Stimuli described as *incongruous-resolved* present an element which is infrequent or incongruous within its context; these elements are more difficult to integrate into the ongoing structure, as they depart from the norm, but do not preclude the processing of the stimulus as a meaningfully structured unit (i.e. a chord sequence in which one of the chords is borrowed from a closely related key, or a sentence which presents a low-cloze probability word but still makes sense).

Stimuli described as *incongruous-unresolved* present an element which cannot be integrated into the unfolding structure of the stimulus (i.e. a chord sequence with a sudden modulation to an unrelated key, or a sentence containing two main verbs).

To maintain maximal comparability between the stimuli in each *congruence* condition, the stimuli all follow the format illustrated in Figure 3.2, namely:

1. beginning of stimulus in *incongruous-resolved* or *normal* condition
2. region of interest (incongruity plus two syllables or crotchets either side)
3. end of stimulus in *incongruous-resolved* or of *normal* condition

This ensures that the beginnings and ends of the stimuli within each set occur in two different *congruence* conditions.

The region of interest (illustrated by the light blue square in Figure 3.2) contains the element which is incongruous in the *incongruous-* conditions and two syllables or crotchets either side of it. This region was kept identical across conditions, using the same piece of sound recording in all three stimuli of any given stimulus set. This allowed not only the control of coarticulation artefacts, but also the study of responses to both the onset and offset of the incongruity.

The three *incongruity types* created for this research project are tied to the notion of *ease of integration* (Hale, 2006), which has been associated with processing difficulty in both music and language (Patel et al., 1998; Patel, 2003; Jurasfky, 2003; Hale, 2006). In the syntactic stimuli, this can be equated with the ease of the transformation necessary to retrieve the meaning of the sentence with a usual sentence structure. For the harmonic stimuli, adopting the paradigm used by Patel et al.'s research team (Patel et al., 1998; Patel, 2003), this was defined as the number of steps separating the key from which the incongruous chord was borrowed from the key of the piece. In the semantic stimuli, the distinction was made by increasing the semantic distance which the incongruity must travel before being applied to the topic (direct association, metaphorical adjective or metaphorical verb). In the rhythmic stimuli, the deciding factor was the extent to which the main beat and phrase structure of the piece needed to be accommodated to integrate the incongruous element. This level of complexity was added to replicate the manipulation which previous authors have made in the main *congruence* condition in previous experiments (Patel et al., 1998), to allow for the assessment of the assumption that more theoretically odd elements are considered more odd by participants (see chapter 4).

This design will become clearer with the examples given in the following sections, which explain how this was implemented across *harmony*, *rhythm*, *semantics* and *syntax*. Example stimuli can be heard at www.carafeatherstone.co.uk/research/stimuli and the full stimulus set is included on the CD provided with this thesis (Appendix B.2).

3.2.1 Music stimuli

Previous studies investigating the processing of music incongruities have used stimuli based on the music of J.S.Bach (Steinbeis et al., 2006; Patel, 2007; Koelsch, Gunter, et al., 2002; Koelsch et al., 2008) considered to be the composer whose work implicitly set the rules of western tonal classical music (Grout, 1973). The stimuli for the present research were based on music by The Beatles. This choice was motivated by the fact that just as Bach's music created the transition from the baroque era to the classical era by building upon and developing previously established musical patterns (Grout, 1973) the music of the Beatles is generally considered to have revolutionized modern western tonal music, bending both traditional harmonic and rhythmic rules to create music which, while being innovative, remained accessible and mainstream. In the words of Robert Greenfield, "People are still looking at Picasso. People are still looking at artists who broke through the constraints of their time period to come up with something that was unique and original. In the form that they worked in, in the form of popular music, no one will ever be more revolutionary, more creative and more distinctive than the Beatles were" (Gross, 2009).

Once musical extracts by The Beatles containing rhythmic or harmonic incongruities had been identified, they were systematically manipulated to form the three *congruence* conditions described above. Each stimulus was created from a different extract, though some songs provided material for several stimuli (the table in Appendix B.1 presents each stimulus triplet with the song from which it was adapted). In order to control for any effects of key, some original extracts were transposed to ensure that all twelve possible keys were equally represented in the stimuli. All stimuli were created using the music software *Sibelius 5* using a standard piano sound from the software's in-built *KontaktPlayer2*. In order to ensure greater control over potential confounds, the dimension which was not the one in which the incongruity occurred (i.e. *rhythm* for the *harmony stimuli* and *harmony* for the *rhythm stimuli*) was varied in a systematic fashion across the stimuli. For instance, if the musical extract presented a strange chord, the chord sequence in which that strange chord occurred was isolated from its original rhythm and re-written with one of four possible rhythmic patterns (rhythmic *frames*). If on the other hand the incongruity was rhythmic, the rhythmic sequence which contained the incongruity was rewritten with one of four possible harmonic patterns (harmonic *frames*). This ensured that there was systematic variance between stimulus triplets and that none of the stimuli were recognizable as extracts from songs by the Beatles, while also ensuring that the stimulus set designed to address incongruity-based aesthetics was built on naturally occurring aesthetically-purposed incongruities.

3.2.1.1 Harmonic stimuli

In these stimuli, the dimension of interest was the harmony, and the rhythm was considered incidental with regards to *congruence*. Thirty six musical extracts which contained harmonic incongruities were identified. These incongruities were further divided into three *incongruity types*, which departed to differing extents from the traditional rules of classical western harmony:

a. Chromatic chord (Neapolitan sixth, diminished chord, minor chord on the 4th degree)

b. Chord borrowed from a closely related key (generally from the key of the dominant)

c. Chord borrowed from an unrelated or loosely related key (major version of the minor relative or key a tone below the original key)

Four rhythmic sequences containing no irregularities were chosen from songs by The Beatles as a basis for creating all the stimuli which presented harmonic incongruities. These rhythmic patterns were chosen so as to vary along two dimensions, namely the type of beat (subdivided into 2 quavers in 4/4 or subdivided into 3 quavers in 6/8) and where most of the rhythmic feel was gained from (the higher notes or the lower notes). The four rhythmic frames, equally represented in the harmonic stimuli, were derived from the following songs:

1. 4/4, higher notes - *Good night*
2. 4/4, lower notes - *Hey Jude*
3. 6/8, higher notes - *Hide your love away*
4. 6/8, lower notes - *With a little help*

An additional four bars were tagged onto the end of the 6/8 sequences, to create an equal number of quavers across all harmonic stimuli. At a tempo of 200 quavers per minute, all sequences last exactly 20s. Three tempi (70, 100 or 130 beats per minute) were equally represented across all music stimuli. Once more, this manipulation ensured systematicity and variance across stimuli (an equal number of quavers in all music stimuli but different tempi and different rhythmic patterns), while allowing the study of the effect of tempo on incongruity processing.

The manipulations outlined above resulted in twelve possible combinations of *incongruity type* (*a*, *b* or *c*) and *rhythmic frame*: three types of incongruities which can be placed upon four different rhythmic patterns. The frames were distributed evenly across the 36 stimulus sets, creating three sets for each combination. For instance, the rhythm based upon Hey Jude was used to create three incongruities of type *a*, three of type *b* and three of type *c*.

To control any effect of tonality (or *key*), some extracts were transposed to ensure that the twelve keys of Western tonal harmony were equally represented. This equal representation was implemented by making sure that each type of harmonic incongruity (a, b, c) had one stimulus set starting in each key in the *incongruous-* conditions (the *normal* stimuli were in the key of the region of interest).

Finally, the 36 *incongruous-resolved* stimuli were manipulated to create the three *congruence* conditions as follows (illustrated in Figure 3.3):

Incongruous-resolved The incongruous chord is followed by a resolution into the key of the stimulus (e.g. a sequence in D with a chord borrowed from C continues in D once the incongruity is resolved).

Normal The key of the stimulus is changed to match the key of the chord of interest so that it no longer sounds odd (e.g. the whole stimulus is in C).

Incongruous-unresolved The extract is the same as in the first *incongruous-resolved* condition up to the incongruous chord, at which point the rest of the stimulus

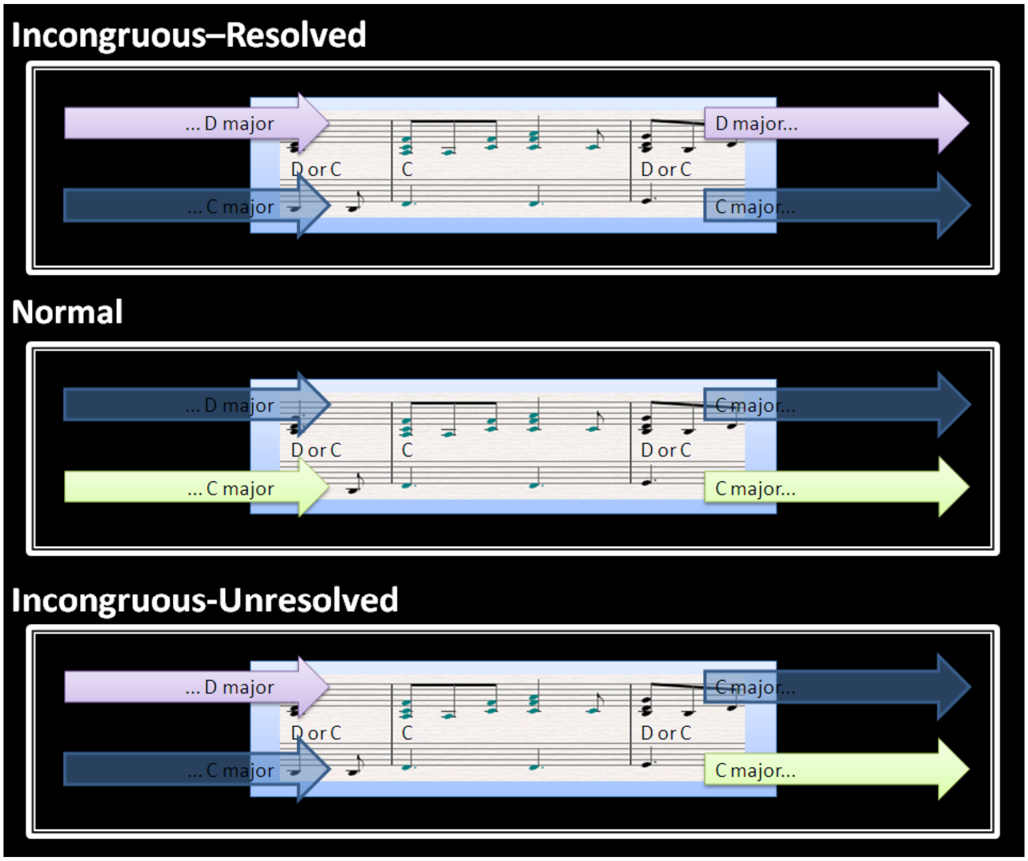

Figure 3.3: Format of all *harmony* stimuli

continues in the key determined by the incongruity (e.g. a sequence in D with a chord borrowed from C continues in C after the region of interest).

Chords leading to and from the region of interest in the *normal* condition were chosen with reference to the Table of Usual Root Progressions provided in Piston's (1978) *Harmony* to ensure that no chord was considered incongruous within the new key.

3.2.1.2 Rhythmic stimuli

Rhythm is understood with reference to a metric frame. The time signature of a piece (e.g. 3/4 or 4/4) determines which beats are strong and which are weak (e.g. respectively 1 (um - pah - pah and 1 and 3 (um - pah - um - pah)). Composers tend to manipulate the emphasis within the bar to create rhythmic tension against the solid background of the metric frame. The tension is often most felt when participants are asked to try to keep track of the original rhythm during a rhythmically ambiguous sequence (Vuust et al., 2006). This metric frame acts as an implicit reference for a musical piece as a whole and is well perceived by non musicians as well as musicians as a solid constant in the piece. The manipulation of the interplay of music with this constant leads to aesthetic and affective effects (Vuust et al., 2006) as well as to incongruity-type responses (Besson et al., 1994).

In these stimuli, the dimension of interest was the rhythm, and the harmony was considered incidental with regards to the three *congruence* conditions. Thirty-six musical extracts which contained rhythmic incongruities were identified. As in the case of the harmonic incongruities, these incongruities were further divided into three types:

a. A change of emphasis in the beats (e.g. making 6/8 sounds like 3/4 but the underlying beat can still be maintained) - the original meter can be maintained

b. A change in the subdivision of beats (e.g. from 2/4 to 6/8) - the original meter can be maintained

c. A change in the number of beats in a bar or a phrase (e.g. having a bar of 5/4 in the middle of a 4/4 piece) - the original meter must be suspended for the addition or removal of a beat and then resume

In order to ensure that no effect of harmony confounded the results with regards to the judgements of how odd a musical extract was, the underlying harmony for each stimulus was systematically manipulated. Four ordinary harmonic sequences from songs by The Beatles were used as a basis for creating all the stimuli which presented rhythmic incongruities. These harmonic patterns were also chosen so as to vary along two dimensions, namely the mode of the extract (major or minor) and the type of musical writing (vertical, emphasising chord progressions or horizontal, emphasising tunes). The harmonic frames were derived from the following songs:

1. Major, vertical - *Hey Jude*

2. Major, horizontal - *All you need is love*

3. Minor, vertical - *I me mine*

4. Minor, horizontal - *While my guitar*

These were made unrecognisable by changing the tune carried by the chords in the original versions of these songs to one which would carry the rhythmic pattern of interest.

As with the harmonically odd stimuli, care was taken to ensure that all stimuli were approximately the same length, manipulating the number of whole bars while maintaining musical phrase closure to make up for differences in time signature. Twelve possible combinations of *incongruity type* and *harmonic frame* were created, resulting in 36 stimulus sets by creating three sets in each combination. Both to control for any effects of key and to lessen the repetitive nature of the task, the stimuli were transposed to ensure each type of rhythmic incongruity (a, b, c) featured one stimulus set in each key (since the manipulation to create the *normal* condition did not affect the key of the stimulus, all stimuli within one set were in the same key).

Finally, the 36 *incongruous-resolved* stimuli were manipulated to create the three *congruence* conditions as follows (illustrated in Figure 3.4):

Unexpected-resolved The incongruous rhythmic pattern occurs after a number of bars containing a regular beat and resolves into the rhythmic pattern of the beginning of the stimulus (e.g. a sequence in 4/4 contains a bar in which the beat emphasis makes it sound like 9/8 before returning to a standard 4/4 pattern).

Normal The beat is adapted to fit with the element of interest, which goes unnoticed (e.g. the whole sequence is in 9/8).

Unexpected-unresolved The stimulus starts the same as in the first *incongruous-resolved* condition but continues with the beat dictated by the incongruous pattern once this pattern has occurred (e.g. a sequence in 4/4 continues in 9/8 after the occurrence of the incongruity).

The difficulty in analysing *rhythm* effects lies in the necessary interplay between the incongruous section and the original meter. Since it is not possible to determine what

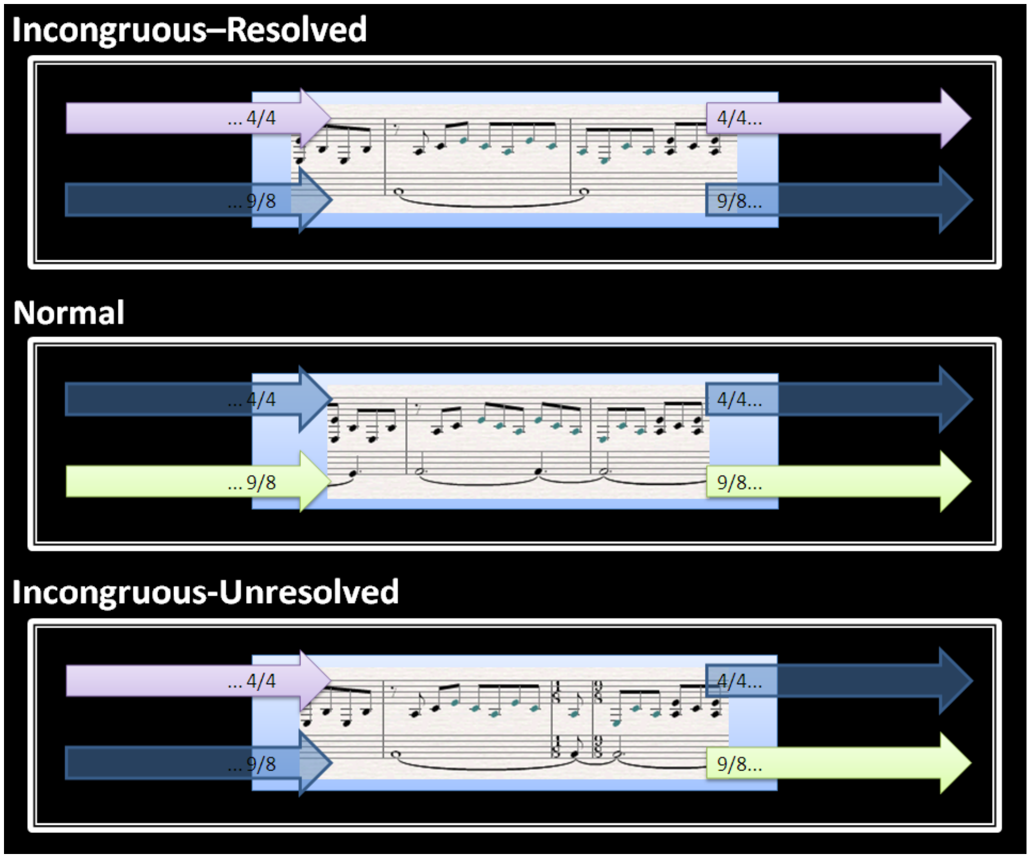

Figure 3.4: Format of all *rhythm* stimuli

meter participants are keeping without requiring them to tap along to the main beat, the analysis of what participants would perceive at the point of resolution is problematic. Nonetheless, these stimuli were constructed with equivalence to the *harmony*, *semantics* and *syntax*, and therefore provide a valid contribution to the effect of incongruities in the processing of music and language.

3.2.2 Language stimuli

To create an equivalence with the music stimuli in the initial selection of incongruities to build upon, the types of language incongruities used in this study were chosen with the following criteria in mind. Firstly, the incongruities needed to show a good fit with the integration paradigm adopted for the music stimuli and throughout the literature investigating parallels between music and language cognition. Secondly, they needed to possess a previously established experimental validity in ERP incongruity studies, so that the patterns elicited by the language incongruities could be compared to those arising from the processing of harmonic and rhythmic incongruities. Finally, the incongruities needed to have been used for effect in naturally occurring language to allow the investigation of the role of incongruities in the eliciting of affective and aesthetic responses.

A survey of the literature reporting an N400 or a P600 associated with semantic or syntactic violations was carried out. The aim was to establish which incongruous semantic and syntactic elements had been associated both with these ERP effects and with a subjective response which resembled those described by Meyer (1956) and Ortony (1975). Only the studies which reported both these aspects of language incongruities are mentioned below. Stimulus creation methods were also observed. Most studies reported the use of stimuli in which the region of interest in the sentence (the incongruous or normal element) varied across conditions. Pynte et al. (1996) created nonsensical stimuli by swapping the ends of sentence between stimuli and therefore varying the semantic content in which the end of the sentence occurred. This achieved an identical element of interest across conditions while manipulating the level of incongruity of the sentence. This method provided an example of how to apply the procedure used for creating the music stimuli to the creation of language stimuli: sentences were created so that the incongruous element was resolved or rendered meaningless by the continuation of the sentence.

3.2.2.1 Semantics stimuli

Metaphors, which Ortony (1975) described as using incongruities to heighten the evocative power of words, were consistently associated with an N400 of amplitude higher than literal sentences and lower than nonsensical sentences (Bonnaud, Gil, & Ingrand, 2002; Coulson & Van Petten, 2002, 2007; Kazmerski, Blasko, & Dessalegn, 2003; Pynte et al., 1996; Tartter et al., 2002). Metaphors therefore seemed the obvious choice for sentences containing semantic violations. Considering the equivalence between notes or chords (or a single beat) and words, sentences were constructed containing incongruities of only one word.

The sentences in the *incongruous-resolved* condition were adapted from metaphors found in song lyrics, novels and poetry. The continuation of the sentence (after the metaphorical word) rendered the metaphor meaningful in the *incongruous-resolved* stimuli and nonsensical in the *incongruous-unresolved* stimuli. The nonsensical sentences were created by continuing the sentence with something relating literally to the metaphorical word (e.g. a sentence about someone's mind *clanking* was ended with a reference to old car engines). In view of potential overlapping of effects of syntactic and semantic incongruities, the semantic sentences presented a simple syntactic structure. Anaphora were avoided as much as possible and care was taken to ensure that no conflict between anaphora and referents affected the processing of the sentences.

As with the music stimuli, three types of incongruity were created. These were designed to necessitate increasing levels of semantic elaboration in order to make sense of the metaphorical sentence:

a. Simple metaphor: A straight-forward *A is B* metaphor (*Lucy's friendship group was an* ***oasis.***)

b. Metaphorical through adjective: A sentence in which the subject is qualified by an adjective which cannot be applied literally to it (*His personality was very* ***spiky.***)

c. Metaphorical through verb: A sentence in which the main verb cannot be applied literally to the subject (*A smile was* ***sprouting*** *from the corners of her mouth.*)

Twelve metaphors of each incongruity type were created. In the resulting 36 *incongruous-unresolved* stimuli, the metaphorical elements were one (non compound) word long. At least two syllables were kept identical either side of the metaphorical element across all three *congruence* conditions.

The three *congruence* conditions were created as follows (illustrated in Figure 3.5):

Figure 3.5: Format of all *semantics* stimuli

Incongruous-resolved The sentence contains a metaphorical element which is given greater meaning in the sentence after the point of occurrence of the metaphorical element (e.g. *Harold's bitterness was slowly **eating** up all the fond memories he had of Belinda.*)

Normal The sentence contains the same element but its use is literal (e.g. *The huge slugs were slowly **eating** up all the plants in Belinda's garden.*)

Incongruous-unresolved The sentence is made up of the beginning of the *original* sentence and the end of the *normal* sentence, creating a sentence which is non-sensical because the metaphorical departure from semantic transparency is not resolved (e.g. *Harold's bitterness was slowly **eating** up all the plants in Belinda's garden.*)

3.2.2.2 Syntax stimuli

Syntactic incongruities used for effect were more difficult to find in the ERP literature. However, a recent study by Thierry et al. (2008) found that the functional shift (the use of a noun in place of a verb), which has been used for effect since the days of Shakespeare, is systematically associated with a LAN and both early and late positivities and not with an N400 effect as could be expected for a linguistic device which emphasises meaning (Lee & Federmeier, 2006; Thierry et al., 2008). Considering the fit of the functional shift with both the sought ERP effect and aesthetic effects, this type of incongruity was an obvious choice for the syntactic stimuli.

The syntactic-shift verbs were chosen and adapted from the review of nouns used as verbs carried out by Clark and Clark (1979) and from a study by Thierry et al. (2008). The functional-shifted nouns retained for this project were those whose use as verbs has not passed into natural language (e.g. *to* ***butter*** *the bread* was rejected, but *to* ***tomato*** *a singer* was retained). These noun-verbs also needed to be semantically transparent so as not to confound a semantic incongruity with the desired syntactic incongruity. Extra care was also taken to ensure that the sentences presented no other syntactically difficult elements such as difficult anaphoric resolutions, subjacency violations or embedded relative clauses, all shown to elicit specific ERP effects (Kluender & Kutas, 1993; Koornneef & Van Berkum, 2006; Patel et al., 1998).

Since the experimental manipulation resulted in a shift in phrase boundaries in *syntax*, clause-final ERP effects were investigated. A literature search showed that the Closure Positivity Shift is prosody-specific (Steinhauer, Alter, & Friederici, 1999; Steinhauer, 2003; Kerkhofs, Vonk, Schriefers, & Chwilla, 2008) and is found at intonational phrase boundaries signaled by pre-constituent lengthening (Isel, Alter, & Friederici, 2005). As well as occurring in spoken meaningful sentences, this effect has been noted in sentences devoid of semantic content (Pannekamp, Toepel, Alter, Hahne, & Friederici, 2005) and in music in which phrase boundaries have been artificially emphasised by extending pauses or prolonging pre-boundary notes (Nan, Knosche, & Friederici, 2006;

Neuhaus, Knosche, & Friederici, 2006). Since the same piece of sound recording was used across all conditions, the prosody of the region of interest of the sentences (the function-shifted word and the two syllables either side of it) was kept constant across conditions. This ensured that differences in prosody could not account for positivities elicited by said area. Concerns were raised over whether this would result in the prosody sounding odd for certain sentence structures, but this concern was eliminated by the pre-norming study auditory controls and the norming study itself.

The three levels of incongruity in *syntax* were decided through consideration of the extent to which the surface structure needed to be modified to retrieve the ordinary way of communicating the same semantic content:

a. Locatum verb: the parent noun is in the objective case in clauses that describe the location of one thing with respect to another (*Jake puts **sugar** in his tea* becomes *Jake **sugars** his tea*)

b. Location verb: the parent noun is in the locative case in clauses that describe the location of one thing with respect to another (*Jake puts the sugar **in the tea*** becomes *Jake **teas** the sugar*)

c. Instrument verb: the parent noun denotes an instrument with which the action is undertaken (*Jake puts sugar in his tea **with the help of a teaspoon*** becomes *Jake **teaspoons** sugar into his tea*)

Twelve sentences were created for each incongruity type, and sentences were adapted as follows to make up the three *congruence* conditions, maintaining as much semantic clarity as possible (illustrated in Figure 3.6).

Incongruous-resolved The sentence uses a functional shift and resolves to a normal structure (e.g. *Because her two-year-old never stops kicking, Mrs Matthew **trousers** her son with great difficulty every morning.*)

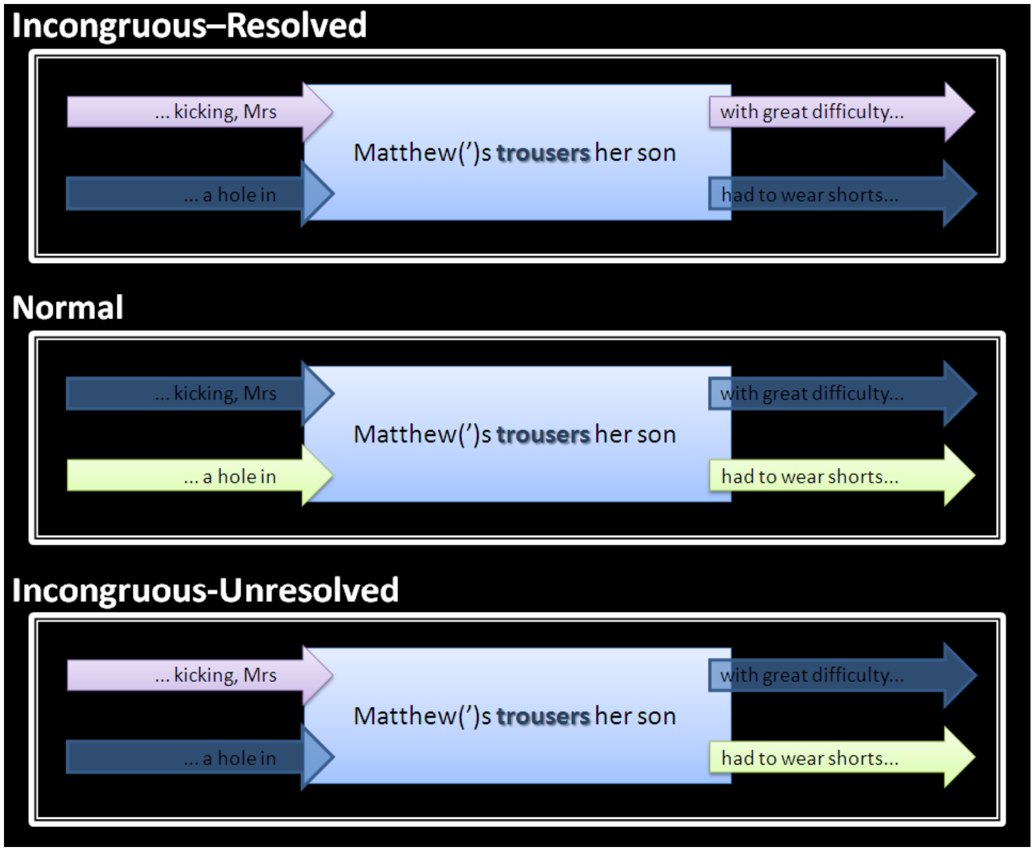

Figure 3.6: Format of all *syntax* stimuli

Normal The sentence uses the functional shift noun words in the *original* as a noun in a sentence with a simple syntactic structure (e.g. *Because Mary had ironed a hole in Matthew's* ***trousers*** *her son had to wear shorts to school.*)

Incongruous-unresolved The sentence is made up of the beginning of the *original* sentence and the end of the *normal* sentence, creating a sentence with an erroneous syntactic structure (e.g. *Because her two-year-old never stops kicking, Mrs Matthew* ***trousers*** *her son had to wear shorts to school.*)

The fact that the region of interest has a different spelling in the *normal* condition was of no consequence since the stimuli were to be auditorily presented.

3.2.2.3 Extra controls for the language stimuli

The sentences designed through this process were first of all assessed in written form by three independent observers for *congruence* condition fitness (see Appendix B.1 for the full list of sentences). These observers were asked to report any sentences which were not abnormal in the *incongruous-unresolved* condition, any sentences which were not completely normal in the *normal* condition, and any sentence in which the incongruity was either too great for the sentence to make sense or too small to be noticed in the *incongruous-unresolved* condition. Sentences were adapted based on feedback before being recorded. Such amendments included minor changes to the structure of some sentences or the content words used to express the meaning of the sentence to ensure that the level of incongruity and comprehension difficulty were affected by the element of interest alone.

All sentences, recorded and edited using Digidesign ProTools7 LE, were spoken by the same female speaker in a neutral tone, avoiding any prosodic emphasis. The use of this professional software allowed the seamless juxtaposition of parts of sentences from different recordings. It was therefore possible to use exactly the same sound section for the region of interest (incongruity + 2 syllables either side) across the three *congruence* conditions of each sentence set, ensuring that no differences in prosody between conditions could affect the results. After editing, all sentence sounded as if they had been spoken as a whole, and flowed at the pace of natural speech.

Once the sentences had been edited, they were submitted to five independent observers to be checked for clarity. A small number of sentences were re-recorded to improve their comprehensibility or to ensure the pace at which they were read was consistent across all the language stimuli.

3.3 Conclusion

This method of creating stimuli resulted in 36 stimulus triplets for level of *hrss* (*harmony, rhythm, semantics* and *syntax*) with three kinds of original incongruity within each level. Within each stimulus triplet, the region of interest (which contains the incongruous element in the *incongruous-resolved* and *incongruous-unresolved* conditions) was identical across all *congruence* conditions. The context of this identical region varies to create the three different levels of *congruence*. Though no direct equivalence between specific elements of music and language is claimed (e.g. a note is equivalent to a word, a bar is equivalent to a noun phrase), the process described thus far has produced computationally equivalent incongruities across four different types of stimuli (*hrss*) by applying the same experimental manipulation of the degree of fit of incongruities with their linguistic or musical context to aesthetically purposed incongruities while maintaining musical or linguistic good form. Since this research is concerned with the effect of this manipulation (i.e. making an element more or less incongruous with its context), this manipulation resulted in a stimulus set which was theoretically fit for the study of aesthetic incongruities across the four components of music and language (harmony, rhythm, semantics, syntax).

Chapter 4

Perceiving theoretical oddity: an initial rating study with the purpose-built stimulus set

4.1 Introduction

This chapter describes a rating experiment (Experiment 1) using the stimulus set described in Chapter 3. Its aim was two-fold: first, to investigate whether theoretically incongruous elements are perceived as odd by listeners regardless of levels of musical training, testing an assumption implicit in previous investigations in this field of research, and second, to provide data with which to select stimuli for future studies based on the *congruence* condition fitness of each stimulus in each stimulus triplet (a stimulus triplet consists of one *normal*, one *incongruous-resolved* and one *incongruous-unresolved* stimulus containing the same target element - see Figure 3.1 in Chapter 3).

4.1.1 Experiment 1

Participants were presented with an equal number of stimuli in each *congruence* condition and *hrss* condition and asked to rate firstly how odd they found the stimulus and secondly how positive or negative their response to the stimulus was. Considering the evidence that mere exposure to a certain grammar enables participants to build up enough knowledge to discriminate between correct and incorrect sequences (Tillmann et al., 2000), it was firstly hypothesised that theoretically incongruous stimuli would receive higher scores on the oddity scale. In accordance with recent work in the psychology of music and language aesthetics (Juslin & Västfjäll, 2008; Thierry et al., 2008) it was secondly hypothesised that aesthetic-affective ratings would show a more positive response to stimuli in the *incongruous-resolved* condition than to the stimuli in the two other conditions. Data were analysed in a 3 (*congruence*) x 4 (*hrss*) design to investigate the effect of incongruities across all four components of music and language, for oddity ratings and aesthetic-affective responses.

Further analysis was concerned with the effect for *incongruity-type* within the *incongruous-resolved* condition. Inoongruity types *a*, *b* and *c* were designed with incremental theoretical processing difficulty, based upon the amount of effort necessary to integrate the incongruity into the ongoing stimulus (see chapter 3 for greater clarification). This comparison was prompted by the fact that different research teams investigating the processing of musical incongruities use this incongruity type distinction as a *congruence* condition in *harmony* stimuli. Koelsch and his colleagues have typically used chromatic chords similar to those used in *incongruity type a* as mildly incongruous chords, and those in *incongruity type c* as strongly incongruous chords. This research team have reported late negative ERP components in the processing of harmonic incongruities. Patel and his colleagues on the other hand have typically used chords borrowed from a nearby key (similar to *incongruity type b*) as mildly incongruous chords, and chords from a distant key (*incongruity type c*) as strongly incongruous chords. This research team have reported late positive ERP components. This *incongruity type* analysis therefore served to investigate any differences between these incongruity types which could account for differences between these two research teams' results.

At the outset of the experiment, participants were asked how many years of musical training they had received. This measure of musical training was chosen, not because of its reliability in terms of determining how intrinsically musical each participant was, but because it would serve as an indicator of whether participants had received, and learnt to manipulate, explicit knowledge regarding the rules of classical harmony. A second aspect of this analysis was to use these data to investigate any effects of musical training on the measures of incongruity processing used in this experiment.

In sum, the data from Experiment 1 were used to address the following questions:

1. What is the effect of theoretical incongruities on oddity responses?

2. Do different kinds of theoretical incongruity have a different effect on responses?

3. What is the effect of theoretical incongruities on aesthetic-affective responses?

4. Is there any effect of musical training on these responses?

4.2 Method

4.2.1 Participants

Thirty six native British-English speakers (mean age = 27.50 years, 14 male) took part in all four rating tasks in this study. This sample size, chosen to match the number of stimulus triplets in each level of *hrss*, met the requirements for a power of 0.95, an alpha of 0.05 and a medium effect size (Faul et al., 2007, minimum number required, 24). The opportunity sample included musicians and non musicians (average number of years of musical training = 3.86, range 0 to 12, median = 3) recruited from the University of Leeds community via posters, departmental emails and the Institute of Psychological Sciences' participant pool scheme which rewards participants with course credits. None of the participants had any diagnosed hearing or language impairments and were naive to the design of the stimuli.

4.2.2 Materials, design and procedure

The participants were tested using a computer running the software *E-Prime* version 1.2. Closed noise cancelling headphones were used to play the music and the sentences and participants were tested individually in a quiet room. The stimuli were those described in detail in Chapter 3. Each participant heard a total of 144 stimuli (36 *harmony*, 36 *rhythm*, 36 *semantics* and 36 *syntax*), consisting of one from each stimulus set, with an equal number of stimuli in each *congruence* condition (12 *normal*, 12 *incongruous-resolved*, and 12 *incongruous-unresolved* in each level of *hrss*). This Latin square approach was adopted so that each participant heard each target element in only one context, to eliminate the confound of memory-based predictions during listening. During listening, a fixation point was provided in the form of an asterisk at the centre of the screen.

The experiment was carried out in four parts (one for each level of *hrss*) during which participants were tested individually in a quiet room. The four parts were completed in two to four sessions, allowing a short break between blocks should participants choose to do more than one in a single session. In all parts of the experiment, participants were instructed to press the space bar as soon as they heard anything strange while listening to a stimulus. The way the experiment was programmed in E-Prime allowed this space bar press to occur without interfering with the presentation of the auditory stimulus, which was played in full regardless of whether the participant had pressed the space bar.

After hearing each music stimulus, participants were presented with two rating scales. These, presented one after the other, were: “Please use the mouse to indicate how odd you thought the piece sounded on the scale below” with a visual analogue scale ranging from “completely normal” to “very odd”, and “Please use the mouse to indicate your emotional response to the piece on the scale below" with a visual analogue scale ranging from "very negative" to "very positive".

After each language stimulus, three scales were used. The first was identical to the one used in the music stimuli: "Please use the mouse to indicate how odd you thought the sentence sounded on the scale below" ("completely normal" to "very odd"). The second was designed as a means of eliminating from any future studies sentences in which participants' aesthetic-affective response could be attributed to the event described by the sentence rather than to the way in which that event was described: "Please use the mouse to indicate your feelings towards the event described in the sentence on the scale below" ("very negative" to "very positive"). The third was designed to capture participants' aesthetic affective response to the wording of the sentence, the facet of participants' response this project is centered on: "How does the wording of the sentence affect the communication of its meaning?" ("It makes the meaning very hard to understand" — "It conveys the meaning in a straightforward and plain way" — "It strongly emphasises the meaning").

Each part lasted on average 25 minutes (30 minutes for the music stimuli, 20 minutes for the language stimuli). The difference in duration between these experiments is due to the difference in duration of the stimuli themselves (on average 22.17s for the music stimuli and 6.44s for the language stimuli). The order of the two parts within each domain (*music* or *language*) was counterbalanced across participants and the stimuli were presented in a different random order within each part for each participant. Data were collected by the experimental software and analysed using *SPSS* 15.0.

4.3 Results

4.3.1 Space bar data

4.3.1.1 Frequency analysis

The data from the space bar presses were collected to identify whether participants reported an element other than the incongruity as odd. This was done by analysing reaction times in relation to the element of interest (incongruity) within each stimulus, and analysing frequencies across *congruence* conditions. Because of the complexity of the design, the software used to collect participants responses only recorded one space bar press per stimulus. Participants were given this information at the outset of the experiment. The space-bar data therefore only reflect the participants' first indication of an incongruous element within the stimulus.

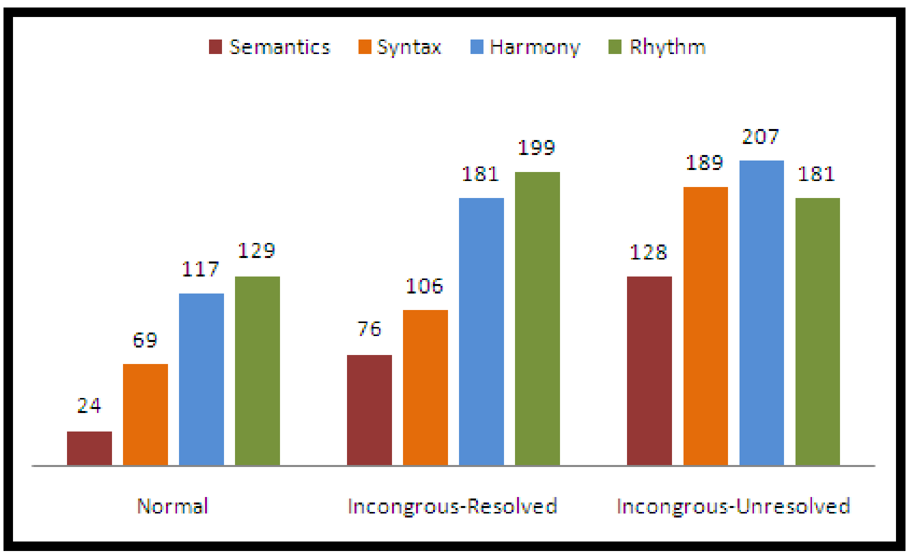

Figure 4.1: Total space bar frequencies by *hrss* and by *congruence* (Max. frequency for each condition cell = 432 (12 x 36))

Figure 4.1 shows the frequencies within each level of *congruence* and each level of *hrss*. Since each participant heard 24 stimuli in which an incongruous element was present (12 *incongruous-resolved* and 12 *incongruous-unresolved* stimuli) in each level of *hrss*, the maximum number of legitimate presses per participant and per *hrss* condition was 24. Thirty-six participants therefore resulted in a maximum 864 presses per level of *hrss* (36 x 24). The lower frequency of space bar presses in the language stimuli compared to the music stimuli can be accounted for by the shorter duration of the language stimuli, requiring participants to respond more rapidly to incongruous elements for the space bar to be recorded. All levels of *hrss* showed an increase in frequency from *normal* to *incongruous-resolved* and from *incongruous-resolved* to *incongruous-unresolved* apart for *rhythm*. The *rhythm* incongruities showed a slight decrease in frequency from *incongruous-resolved* to *incongruous-unresolved*.

A χ^2 goodness of fit test was used to assess the significance of any effect of *congruence* on the distribution of frequencies. A significant effect for *congruence* was seen overall ($\chi^2 = 127.11, df = 2, p < 0.001$) as well as within each level of *hrss* (see Table 4.1).

Table 4.1: Chi-square for goodness of *congruence* fit, overall and by *hrss*

	Chi-square test
Overall	$X^2 = 127.11$ $df = 2$ $p < 0.0001$
Harmony	$X^2 = 25.49$ $df = 2$ $p < 0.0001$
Rhythm	$X^2 = 15.58$ $df = 2$ $p = 0.0004$
Semantics	$X^2 = 71.16$ $df = 2$ $p < 0.0001$
Syntax	$X^2 = 62.65$ $df = 2$ $p < 0.0001$

4.3.1.2 ANCOVA

A by-subjects analysis was conducted, to obtain the total frequency of space bar presses for each subject within each level of *congruence* and *hrss*. The average by-subject frequencies within each *congruence* condition and each level of *hrss*, displayed in Figure 4.2, were analysed in a 3 (*congruence*) x 4 (*hrss*) ANOVA to investigate the effect of *congruence* on mean frequency of space bar presses. Number of years of musical training was added as a covariate to investigate any effects of training in these data.

This 3 x 4 ANOVA revealed a significant main effect for *congruence* ($F(2, 68) = 28.08, p < 0.0001, \eta^2 = 0.45$). A follow-up Bonferroni analysis revealed a significant increase in mean frequency from *normal* to *incongruous-resolved* and from *incongruous-resolved* to *incongruous-unresolved* (both differences $p < 0.001$). A significant main

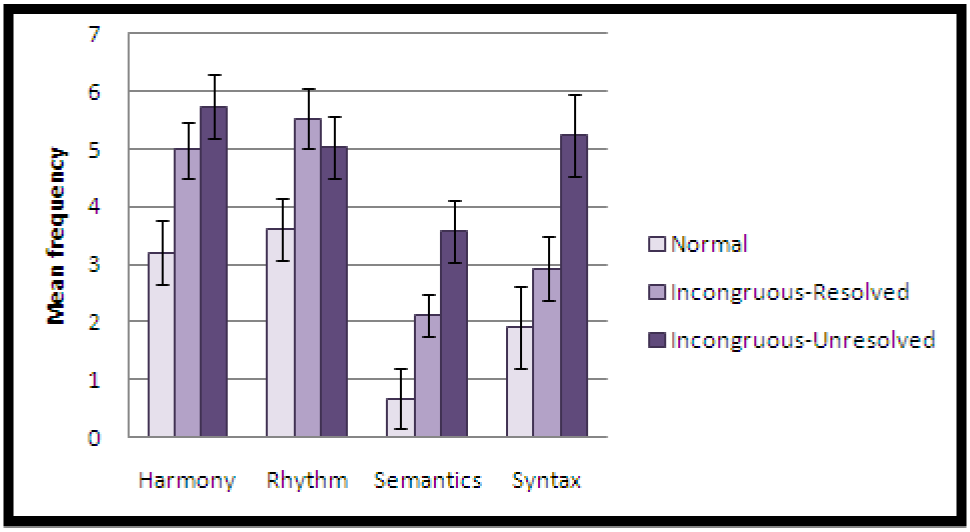

Figure 4.2: Mean space bar frequencies (and standard errors) by subjects by *hrss* and by *congruence*

effect for *hrss* was also found ($F(3, 102) = 4.53, p = 0.005, \eta^2 = 0.12$), with a Bonferroni analysis showing the *semantics* stimuli receiving on average fewer presses than the *harmony* stimuli ($p < 0.001$), *rhythm* stimuli ($p < 0.001$), and *syntax* ($p = 0.04$). The significant *congruence* x *hrss* interaction effect ($F(6, 204) = 2.67, p = 0.02, \eta^2 = 0.07$) called for simple main effects analyses to look at the effect of *congruence* in each of the levels of *hrss*. The results of the simple main effects analyses, reported in Table 4.2, show a significant main effect in all levels of *hrss* bar *rhythm* where the main effect narrowly failed to reach significance. In all levels of *hrss* also *normal* stimuli received on average significantly fewer space bar presses than *incongruous-resolved* and *incongruous-unresolved* stimuli, but only in the language stimuli did *incongruous-unresolved* stimuli receive significantly more presses than *incongruous-resolved* stimuli.

Table 4.2: Simple main effects analysis for the effect of *congruence* on space bar mean frequency within each level of *hrss*

	Main effect	Post hoc analysis
Harmony	$F(2,68) = 13.44, p < 0.001, \eta^2 = 0.28$	n <* i-r < i-u
Rhythm	$F(2,68) = 3.09, p = 0.055, \eta^2 = 0.07$	n <* i-r > i-u
Semantics	$F(2,68) = 12.73, p < 0.001, \eta^2 = 0.27$	n <* i-r <* i-u
Syntax	$F(2,68) = 14.01, p < 0.001, \eta^2 = 0.29$	n <* i-r <* i-u

* indicates a significant difference at p<0.01 in a Bonferroni pairwise comparison

With regards to the covariate, number of years of musical training showed no significant interactions with either *congruence* ($F(3,102) = 1.09, p > 0.05, \eta^2 = 0.01$), *hrss* ($F(3,102) = 1.09, p > 0.05, \eta^2 = 0.01$), or *congruence* x *hrss* ($F(3,102) = 1.09, p > 0.05, \eta^2 = 0.01$). This suggested that the extent of participants's musical training had no significant bearing on the by-subjects space bar data.

4.3.1.3 Summary

The space bar data showed a significant effect for condition, with participants pressing the space bar in significantly more stimuli containing incongruous elements than in *normal* stimuli. In view of the scarcity of these data in the language stimuli, these results must be interpreted with caution. Nonetheless, the space bar data, which had the main purpose of denoting which parts of a stimulus were erroneously unusual, indicated data patterns in line with expected trends and were sufficient for the norming procedure described in Chapter 5.

4.3.2 Rating data

Rating data were generated when participants used the mouse to click at a point along the visual analogue scales described in the Method section. The software calculated the number of pixels on the horizontal line between the zero point on the scale and the place where participants clicked on the screen. This number of pixels was the raw score. As is common practice with visual analogue scales, these raw data were z-transformed to maximise the use of each scale within each participant. Since original scores also showed a discrepancy between the range of rating scores seen in music and in language, z-scores were calculated within each level of *hrss* within each participant. This ensured comparability between the data obtained with music and language stimuli. This approach was adopted for the rest of the analyses of rating data presented in this thesis.

For both the oddity rating and the aesthetic-affective ratings, data were analysed in a 4 (*hrss*: *harmony, rhythm, semantics, syntax*) x 3 (condition: *normal, incongruous-resolved, incongruous-unresolved*) ANCOVA, with years of musical training as a co-variate. Significant interaction effects were followed by simple main effects analyses investigating the main effect for condition on ratings within each level of *hrss*.

4.3.3 Effect of musical training

Table 4.3 displays the elements of the ANCOVA analyses concerned with the effect of musical training on the behavioural outcomes in this experiment. Previous studies in this area have demonstrated effects of musical training on the processing of music (e.g. Patel et al., 1998), an effect which we expect to be replicated throughout this thesis. Studies have also shown effects of musical training on language processing, notably in the areas of verbal memory (Ho, Cheung, & Chan, 2003) and linguistic timing (Overy, 2003). Though the language stimuli used in this project do not resemble those on which such an effect has been shown, leading to low expectations of any effect of musical training on the language outcomes in this research, the beneficial effects of musical training on verbal abilities and other areas of human cognition (Forgeard et al., 2008; Schlaug et al., 2005) called for an analysis of any effect of musical training on language outcomes as well as music outcomes.

Table 4.3: Analyses of the effects of musical training on the rating data

Rating	Analysis		Outcome
Oddity rating	Between-subjects effects	Years of musical training	$F(1,34) = 0.45, p = 0.51, \eta^2 = 0.01$
	Within-subjects effects	Years x Condition	$F(2,68) = 1.84, p = 0.17, \eta^2 = 0.05$
		Years x hrss x Condition	$F(6,204) = 1.44, p = 0.20, \eta^2 = 0.04$
Aesthetic-affective rating (wording)	Between-subjects effects	Years of musical training	$F(1,34) = 0.22, p = 0.46, \eta^2 = 0.02$
	Within-subjects effects	Years x Condition	$F(2,68) = 1.74, p = 0.18, \eta^2 = 0.05$
		Years x hrss x Condition	$F(6,204) = 0.81, p = 0.56, \eta^2 = 0.02$
Affective rating (event - language only)	Between-subjects effects	Years of musical training	$F(1,34) = 1.56, p = 0.22, \eta^2 = 0.04$
	Within-subjects effects	Years x Condition	$F(2,68) = 2.06, p = 0.14, \eta^2 = 0.06$
		Years x hrss x Condition	$F(2,204) = 0.61, p = 0.52, \eta^2 = 0.02$

No significant main effects were found for *years of musical training* as a between-subjects factor, and this factor showed no significant interactions with the other factors in this experiment. Consequently, for the rest of this chapter, the effects of *years of musical training* will not be reported in the analyses. It will be retained as a covariate to control for any subtle effects on the other factors in the analyses.

This lack of significance could have been due to the a posteriori approach adopted here and to the disparity in years of musical training in this sample, notably in the higher end of musical training (see Figure 4.3). More stringent control is necessary to provide greater insight into any effects of musical training on music and language processing. This control was implemented in the experiment which forms the culmination of this thesis (Chapters 6 to 8), by recruiting only participants at the two extremes of the spectrum: complete novices (up to one year of musical training in childhood) and highly trained musicians (having received either Grade 8, in a musical instrument or in voice, and/or an A-level in music).

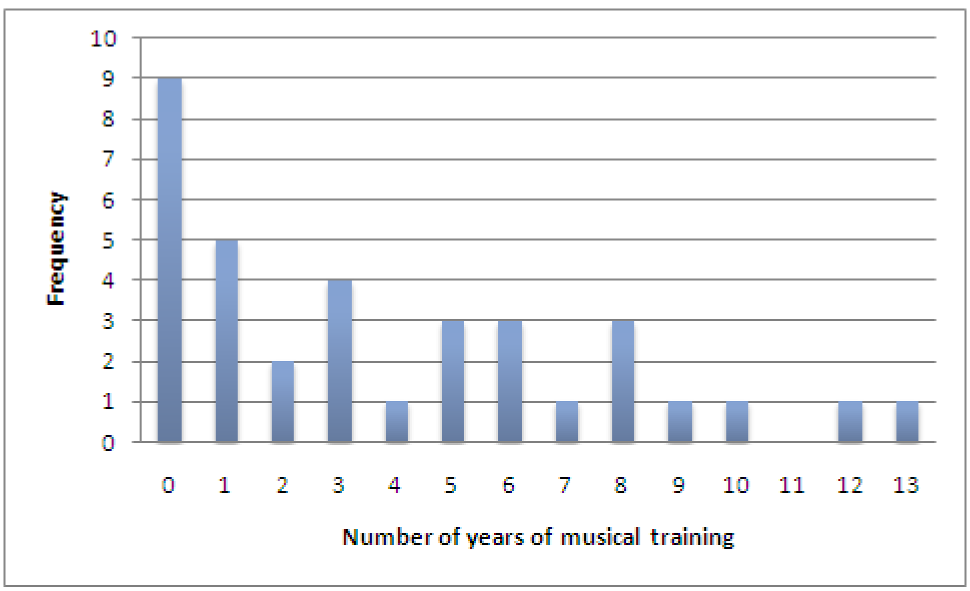

Figure 4.3: Distribution of number of years of musical training in the participant sample from Experiment 1

4.3.4 Oddity ratings

4.3.4.1 ANCOVA

Figure 4.4 presents the mean z-scores for oddity ratings in the four levels of *hrss* in all three *congruence* conditions. A higher score indicated that the participant thought the stimulus was more odd. Mean scores showed an overall increase in oddity rating between the three *congruence* conditions (*normal* < *incongruous-resolved* < *incongruous-unresolved*), concurrent with expectations. However, when splitting the data by *hrss*, this trend is seen to be noticeably stronger in the language stimuli than in the music stimuli. Both types of language stimuli (*semantics* and *syntax*) showed overlapping scores, as did both types of music stimuli (*harmony* and *rhythm*).

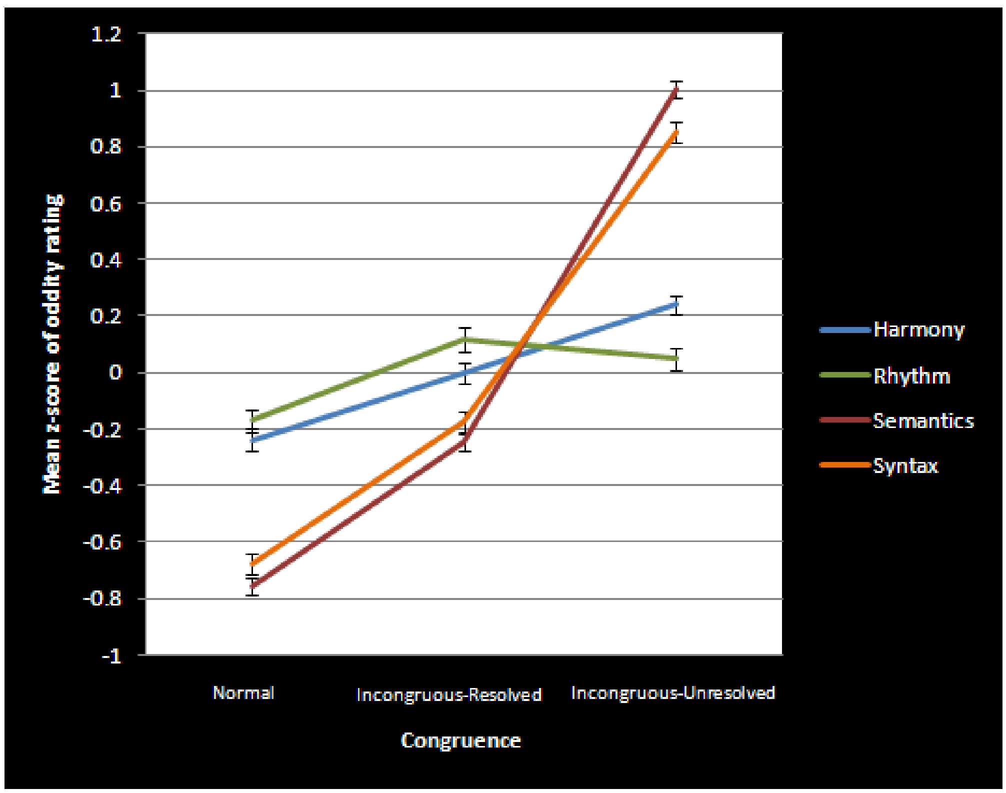

Figure 4.4: Mean oddity z-scores (and standard errors) by *congruence* and by *hrss*

As seen in Table 4.4, the maxima and minima in the raw data were equivalent across all four levels of *hrss*, showing slightly wider ranges in the music stimuli than in the language stimuli. Once data had been averaged by participant and by *congruence* condition, the ranges were smaller for the music stimuli (1.25 and 1.15) than for the language stimuli (2.26 and 2.32).

Table 4.4: Ranges of scores within each level of *hrss*

		Harmony	Rhythm	Semantics	Syntax
Z-scores per participant	Maximum	3.19	3.74	2.80	3.15
	Minimum	-2.31	-2.94	-1.33	-1.98
	Range	5.50	6.68	4.13	5.13
Per participant averaged per condition	Maximum	0.62	0.57	1.25	1.22
	Minimim	-0.63	-0.58	-1.01	-1.10
	Range	1.25	1.15	2.26	2.32

This can be explained by the wider variability shown within each level of *congruence* in the music stimuli than in the language stimuli (see Table 4.5), resulting in less extreme scores once the data points had been averaged within each level of *congruence*.

Table 4.5: Standard deviation of *oddity* z-scores by *congruence* and by *hrss*

	Normal	Incongruous-Resolved	Incongruous-Unresolved	Average
Harmony	0.91	0.98	1.00	0.96
Rhythm	0.97	1.01	0.95	0.98
Semantics	0.47	0.71	0.75	0.64
Syntax	0.79	0.77	0.71	0.76

The ANCOVA carried out on oddity ratings showed a strong main effect for *congruence* was evident ($F(2, 68) = 242.97, p < 0.001, \eta^2 = 0.88$). As expected from the data shown in Figure 4.4, a significant linear trend for *congruence* was found ($F(1, 35) = 736.93, p < 0.001$). A post-hoc Bonferroni analysis revealed that oddity scores increased significantly at each step of the linear trend (*normal* < *incongruous-resolved* < *incongruous-unresolved*, both differences $p < 0.001$).

No significant effect for *hrss* was present ($F(3, 102) = 0.64, p = 0.59, \eta^2 = 0.02$), but a significant *congruence* x *hrss* interaction was noted ($F(6, 204) = 49.30, p < 0.001, \eta^2 = 0.59$), calling for simple main effects analyses to investigate the effect of *congruence* within each level of *hrss*.

4.3.4.2 Simple main effects

As shown in Table 4.6, a significant main effect for *congruence* was found in three of the four levels of *hrss*, being strongest in *semantics*, then in *syntax* and weakest but still of moderate effect size in *harmony*. All three showed a significant increase in mean oddity scores between each of the levels of *congruence* in the Bonferroni post hoc analyses.

Table 4.6: Simple main effect analyses on the *oddity* rating data within each level of *hrss*

	ANCOVA	Post hoc analysis
Harmony	$F(2,68) = 14.27, p < 0.001, \eta^2 = 0.30$	n <* ir <* iu
Rhythm	$F(2,68) = 1.71, p = 0.19, \eta^2 = 0.05$	-
Semantics	$F(2,68) = 347.23, p < 0.001, \eta^2 = 0.91$	n <* ir <** iu
Syntax	$F(2,68) = 167.66, p < 0.001, \eta^2 = 0.83$	n <* ir <** iu

n = normal, ir = incongruous-resolved, iu = incongruous unresolved

* indicates a significant difference at p < 0.005 in a Bonferroni pairwise comparison

** indicates a significant difference at p ≤ 0.001 in a Bonferroni pairwise comparison

When the covariate *years of musical training* was excluded from the analysis, a significant main effect of rhythm was found ($F(2, 70) = 8.37, p = 0.001, \eta^2 = 0.19$), with a significant increase between *normal* and *incongruous-resolved* ($p = 0.002$) but no significant difference between *incongruous-resolved* and *incongruous-unresolved* ($p = 1.00$). However this effect was rendered non-significant when the covariate was included in the analysis. This observation further justifies the importance of strictly controlling musical training in subsequent experiments investigating incongruity processing across domains.

4.3.4.3 Effect of *incongruity-type* in the *incongruous-resolved* stimuli

When investigating the effect of *incongruity-type*, the point of interest for comparison with previous studies lies not in a *incongruity-type* x *congruence* interaction, but in the differences between incongruity types in the stimuli forming the *incongruous-resolved* condition. The degree of theoretical incongruity, based on considerations of syntactic trees, harmonic relationships and rhythmic complexity, increased from *a* to *b* and from

b to c.

An ANCOVA was therefore conducted within each level of *hrss* to ascertain the effect of *incongruity-type* (a, b and c) on oddity ratings in the *incongruous-resolved* condition. In all levels of *hrss*, a significant effect for *incongruity-type* was seen (*harmony* $F(2,70) = 6.02, p = 0.004$, *rhythm* $F(2,70) = 15.16, p < 0.001$, *semantics* $F(2,70) = 17.10, p < 0.001$ and *syntax* $F(2,70) = 3.35, p = 0.05$).

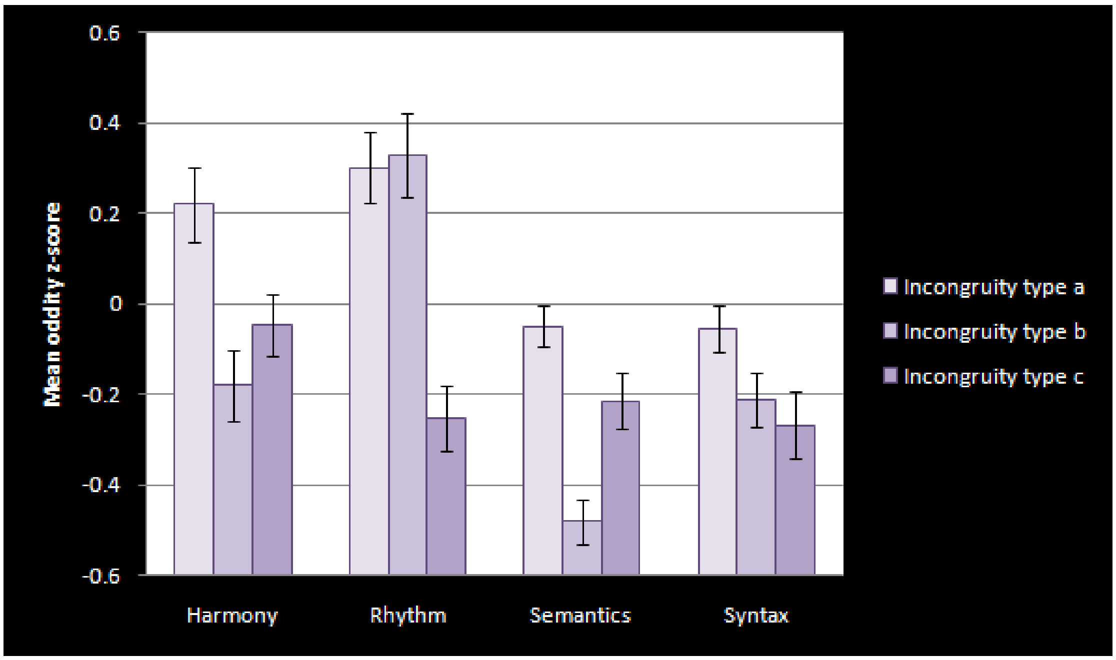

Figure 4.5: Mean *oddity* z-scores (and standard errors) in the *incongruous-resolved* condition by *hrss*

As illustrated in Figure 4.5, in harmony, the only significant difference was that *incongruity-type a* incongruities (chromatic chords, similar to those used by Koelsch and his colleagues) were rated as significantly more odd than *incongruity-type b* incongruities (near-key chord, similar to those used by Patel and his colleagues) ($p = 0.01$). No significant difference was seen between the ratings of near and distant chords (respectively incongruity types b and c, $p = 0.65$) despite a trend in the expected direction ($b < c$). Although Patel and his colleagues usually used the near/distant distinction as a congruence condition criterion, their distant chords were typically more distant than the ones used in this stimulus set. This is accounted for by the fact that the ones used here were all plausibly borrowed from a different key, rather than chosen, with little regard to musical form, from a deliberately very distant key.

A 2 (*congruence*: *normal*, *incongruous-resolved*) x 3 (*incongruity-type*: *a*, *b*, *c*) ANCOVA was conducted to see whether the difference in ratings between the two *congruence* conditions was affected by the *incongruity-type* of the incongruity. No significant *congruence* x *incongruity-type* was found ($F(2,68) = 1.71, p = 0.19, \eta^2 = 0.05$) and neither was a significant three-way interaction with *years of musical training* ($F(2,68) = 0.10, p = 0.90, \eta^2 = 0.003$). These data suggest that despite the differences noted in the one-way ANOVA carried out on the *incongruous-resolved* condition, neither the effect of *incongruity type* nor the effect of musical training impacted on differences between *congruence* conditions.

Another factor distinguishing the stimuli used by these two leading research teams is the placement of the incongruity within the stimuli. This is developed further in chapter 8 which demonstrates that the notion of harmonic closure is able to account for the two components, both elicited by the stimuli created at the outset of this thesis: a late positivity upon the incongruity and a late negativity upon its lack of resolution.

The results from the other levels of *hrss* are displayed in Table 4.7. The closest trend to that observed in the *harmony* stimuli was in the *semantics* stimuli where *incongruity-type b* (a metaphorical adjective) received a significantly lower oddity rating than both incongruity types *a* (a straight metaphor, $p < 0.001$) and c (a metaphorical verb, $p = 0.004$; note, this difference was not significant in the *harmony*). This is the first of many similarities between harmony and semantics (where the other levels of *hrss* differ) to emerge throughout this thesis.

Table 4.7: Effect of *incongruity-type* within *incongruous-resolved* stimuli on *oddity* ratings within each level of *hrss*

	One-way ANCOVA	Post Hoc Analysis
Rhythm	$F(2,68) = 10.35, p < 0.001, \eta^2 = 0.23$	$c <^{**} a, c <^{**} b$
Semantics	$F(2,68) = 14.28, p < 0.001, \eta^2 = 0.30$	$b < a, b <^{*} c$
Syntax	$F(2,68) = 3.08, p = 0.06, \eta^2 = 0.08$	*n.s.*

* indicates a significant difference at p < 0.005 in a Bonferroni pairwise comparison

** indicates a significant difference at p < 0.001 in a Bonferroni pairwise comparison

The *rhythm* stimuli showed a significantly lower rating in *incongruity-type c* (involving a change in the number of beats in a bar) than the other two incongruity types (*a* and *b*, involving a change of emphasis in the main beat or a change of the structure of a beat, $p < 0.001$) and no difference between the ratings of these two incongruity types ($p = 1$). This suggests that rhythmic incongruities in which the number of beats in a bar changes are perceived as less odd than incongruities which affect the structure of the beat or the rhythmic emphasis in the bar. However, this incongruity type prevents the original meter from being maintained, and thus leads to difficulties at the point of resolution. The effect of *subtype* narrowly failed to reach significance in the *syntax* stimuli.

4.3.4.4 Oddity ratings - Summary

The oddity ratings showed that stimuli containing a theoretically incongruous element (whether it resolved or not) received a significantly higher oddity rating than the *normal* stimuli across all four components of music and language, regardless of musical training. Unresolved stimuli received the highest oddity ratings in all levels of *hrss* bar *rhythm*. Possible reasons for this, linked to the notion of meter, are provided in the discussion section of this chapter.

4.3.5 Aesthetic-affective ratings

4.3.5.1 ANCOVA

On both the aesthetic-affective scale and the affect-in-relation-to-event scale, a more negative score indicated a more negative response and a more positive score a more positive response. The zero-point was the middle of the scale, representing a response which was neither negative nor positive in valence. Figure 4.6 shows the mean scores for each *congruence* condition on these two scales designed to capture the aesthetic-affective response to aesthetic incongruities themselves (as opposed to the described event, in the language stimuli).

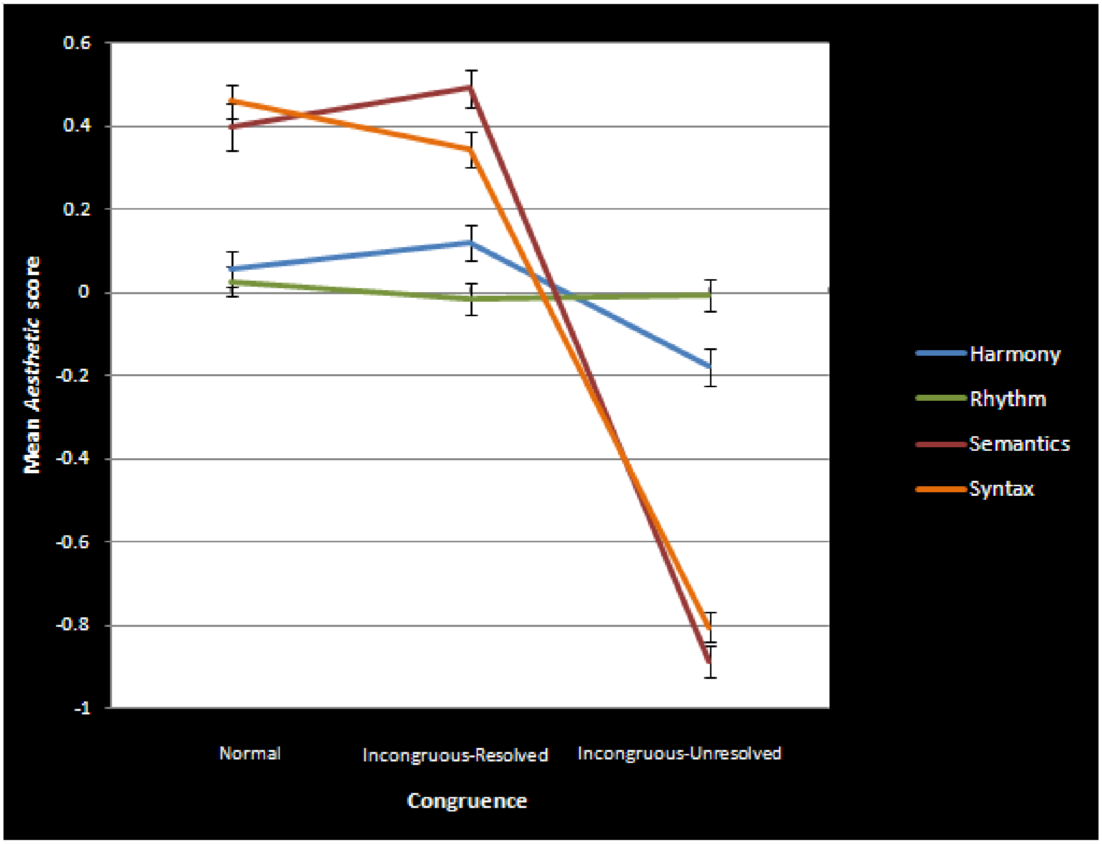

Figure 4.6: Mean *aesthetic-affective* rating z-scores (and standard errors) by *congruence* and by *hrss*

Once more, a wider range of mean scores was seen in the language stimuli than in the music stimuli. Taking into account the maxima, minima and standard deviations (see Tables 4.8 and 4.9), the same conclusion can be drawn as in the case of the oddity ratings, namely that the disparity in scores in the music stimuli led to less extreme mean scores than in the language stimuli once scores were averaged across participants within each level of *congruence*.

Table 4.8: *Affective-aesthetic* rating data: maxima, minima and ranges

	Harmony	Rhythm	Semantics	Syntax
Maximum	3.31	3.34	3.37	3.27
Minimum	-3.35	-2.43	-3.32	-3.10
Range	6.66	5.77	6.69	6.37
Maximum	0.55	0.77	0.97	0.89
Minimim	-0.36	-0.46	-1.26	-1.15
Range	0.91	1.23	2.23	2.04

Table 4.9: *Affective-aesthetic* rating data: standard deviations by *hrss* and by *congruence*

	Normal	Incongruous-Resolved	Incongruous-Unresolved	Average
Harmony	0.98	0.99	0.95	0.97
Rhythm	0.95	1.02	0.98	0.98
Semantics	0.60	0.87	0.78	0.75
Syntax	0.84	0.88	0.69	0.80

In both *harmony* and *semantics*, a slight increase was seen in aesthetic ratings between the *normal* and the *incongruous-resolved* condition, followed by a decrease from *incongruous-resolved* to *incongruous-unresolved*. In the *rhythm* stimuli, little difference was seen between *congruence* conditions. In the *syntax* stimuli, a small decrease in mean ratings from *normal* to *incongruous-resolved* was followed by a large decrease from *incongruous-resolved* to *incongruous-unresolved*.

When all levels of *hrss* were pooled together, the *normal* and the *incongruous-resolved* conditions both had a mean rating of 0.236, whereas the *incongruous-unresolved* condition received a much more negative rating (-0.468). These observations were reflected in the outcome of the 3 (*congruence*) x 4 (*hrss*) ANCOVA, which showed a significant main effect for *congruence* ($F(2, 68) = 158.98, p < 0.001, \eta^2 = 0.82$) following which a Bonferroni post hoc analysis showed that the only significant differences were the *incongruous-unresolved* condition receiving a significantly lower rating than the other two *congruence* conditions (both differences $p < 0.001$). No significant main effect for *hrss* was found ($F(3, 102) = 1.23, p = 0.30, \eta^2 = 0.04$). As with the oddity ratings, a significant *hrss* x *congruence* interaction ($F(6, 204) = 29.16, p < 0.001, \eta^2 = 0.46$) called for simple main effects analyses within each level of *hrss* to explore the effect of *congruence* in a meaningful way.

4.3.5.2 Simple main effects analyses

Table 4.10 displays the outcomes of the simple effects analyses and Bonferroni post hoc analyses for each level of *hrss* on the aesthetic-affective ratings.

Table 4.10: *Affective-aesthetic* rating data: main effect for *congruence* and post hoc anlayses within each level of *hrss*

	ANCOVA	Post hoc analysis
Harmony	$F(2,68) = 6.50$, $p = 0.003$, $\eta^2 = 0.16$	iu <* n, iu <* ir
Rhythm	$F(2,68) = 0.17$, $p = 0.82$, $\eta^2 = 0.01$	-
Semantics	$F(2,68) = 99.56$, $p < 0.001$, $\eta^2 = 0.75$	iu <** n, iu <** ir
Syntax	$F(2,68) = 107.17$, $p < 0.001$, $\eta^2 = 0.76$	iu <** n, iu <** ir

n = normal, ir = incongruous-resolved, iu = incongruous unresolved

* indicates a significant difference at p < 0.02 in a Bonferroni pairwise comparison

** indicates a significant difference at p < 0.001 in a Bonferroni pairwise comparison

All levels of *hrss* bar the *rhythm* stimuli showed a significant main effect for *congruence*. *Harmony*, *semantics* and *syntax* all showed significantly lower ratings in the *incongruous-unresolved* condition than in the other two *congruence* conditions. Both *harmony* and *semantics* showed a slight increase in affective-aesthetic ratings from *normal* to *incongruous* resolved, but this increase was not significant in either of these two levels of *hrss*.

4.3.5.3 Affective response to the event described in the sentence

In the language stimuli only, participants also rated their emotions relating to the event described in the sentence. These data, shown in Figure 4.7 were analysed with the aim of removing from the stimulus set any stimuli which showed scores which stood out from the pool of ratings as an "affective outlier". In a by-items z-score analysis (outliers being identified as stimuli with z-scores greater than 2.5), scores ranged from -1.48 to 1.67. Consequently, no sentences were considered problematic due to the emotional nature of the event described.

The trends shown in these data suggested a main effect for *congruence* rather than an effect of sentence content per se and largely overlapping scores between the data gathered with both types of language stimuli. These data were analysed in a 3 (*congruence*) x 2 (*hrss*:*semantics*,*syntax*) ANCOVA.

In line with these observations, the ANCOVA revealed a significant main effect for

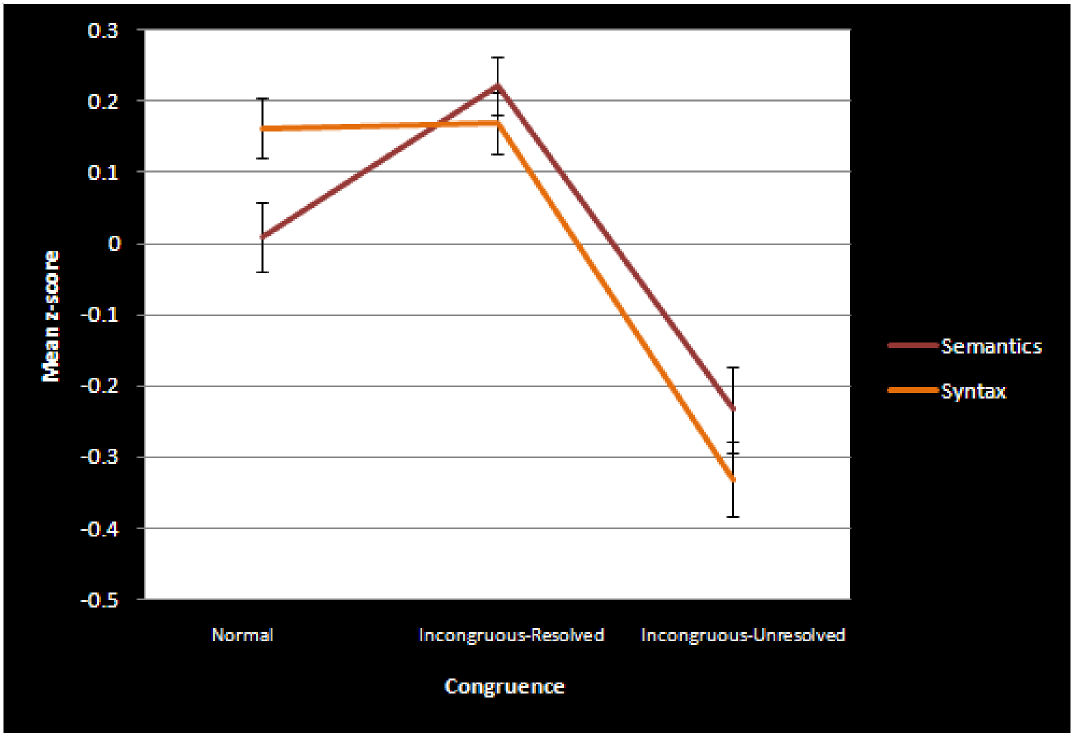

Figure 4.7: Mean z-scores (and standard errors) of the ratings of feelings towards the events described in the sentences by *congruence* and by *hrss*

congruence ($F(2, 68) = 20.85, p < 0.001, \eta^2 = 0.38$). Bonferroni post hoc analyses revealed that the *incongruous-unresolved* condition received a significantly lower mean rating than than the other two *congruence* conditions (both differences $p < 0.001$).

No signficant main effect for *hrss* was found ($F(1, 34) = 0.99, p = 0.33, \eta^2 = 0.03$), confirming the overlap of scores between the two types of stimuli mentioned above. No significant *hrss* x *congruence* interaction was found either ($F(2, 68) = 2.00, p = 0.14, \eta^2 = 0.06$) suggesting that the experimental manipulation had a similar effect on both types of stimuli despite the slight difference between trends in the first two *congruence* conditions (an increase from the *normal* to the *incongruous-resolved* condition in *semantics* but no such increase in *syntax*).

When *years of musical training* was not included as a covariate in the analysis, a significant *hrss* x *congruence* interaction was found (F$(2, 70) = 3.87, p = 0.03, \eta^2 = 0.1$). Simple main effects analyses revealed a significant effect for *congruence* in both *semantics* ($F(2, 70) = 13.87, p < 0.001, \eta^2 = 0.28$) and *syntax* ($F(2, 70) = 25.89, p <$

$0.001, \eta^2 = 0.43$). Bonferroni post hoc analyses revealed that the *incongruous-resolved* condition received a significantly higher rating than the other two *congruence* conditions in *semantics* and that the *incongruous-unresolved* condition received a significantly lower rating than the other two *congruence* conditions in *syntax*.

4.3.5.4 Comparing Aeshtetic-affective ratings and Oddity ratings

Figure 4.8 presents the mean aesthetic ratings for each level of *hrss* and each *congruence* condition (in solid lines) superimposed on the mean oddity ratings for each level of *hrss* and each *congruence* condition (in dotted lines). Each type of rating is given its own vertical axis, with the oddity rating being on the left axis and the aesthetic rating on the right. Comparing the data this way brings out a mirror image of oddity ratings in aesthetic ratings in *syntax* and, most strikingly in *rhythm*. This pattern, which does not seem to occur in the *harmony* or the *semantics* stimuli suggests that the oddity ratings themselves, rather than the incongruities, may have had an effect on the aesthetic-affective ratings *rhythm* and *syntax*.

Based upon these observations, one would expect to see a stronger linear correlation between oddity and aesthetic scores in *rhythm* and *syntax* than in *harmony* and *semantics*. A significant Pearson's product moment correlation was found between aesthetic-affective ratings and oddity ratings in all levels of *hrss* (see Table 4.11). These correlations were in fact of similar magnitude across both stimulus types within each domain. This could be due to the fact that the discrepancy between mean aesthetic scores in *rhythm* was very small, weakening the correlation which looks the strongest when the aesthetic and oddity data are presented on separate scales.

Table 4.11: Correlations between *oddity* and *aesthetic-affective* ratings within each level of *hrss*

	Pearson's product moment correlation		
Harmony	$r = -0.58$	$N = 108$	$p < 0.001$
Rhythm	$r =- 0.36$	$N = 108$	$p < 0.001$
Semantics	$r = -0.87$	$N = 108$	$p < 0.001$
Syntax	$r = -0.92$	$N = 108$	$p < 0.001$

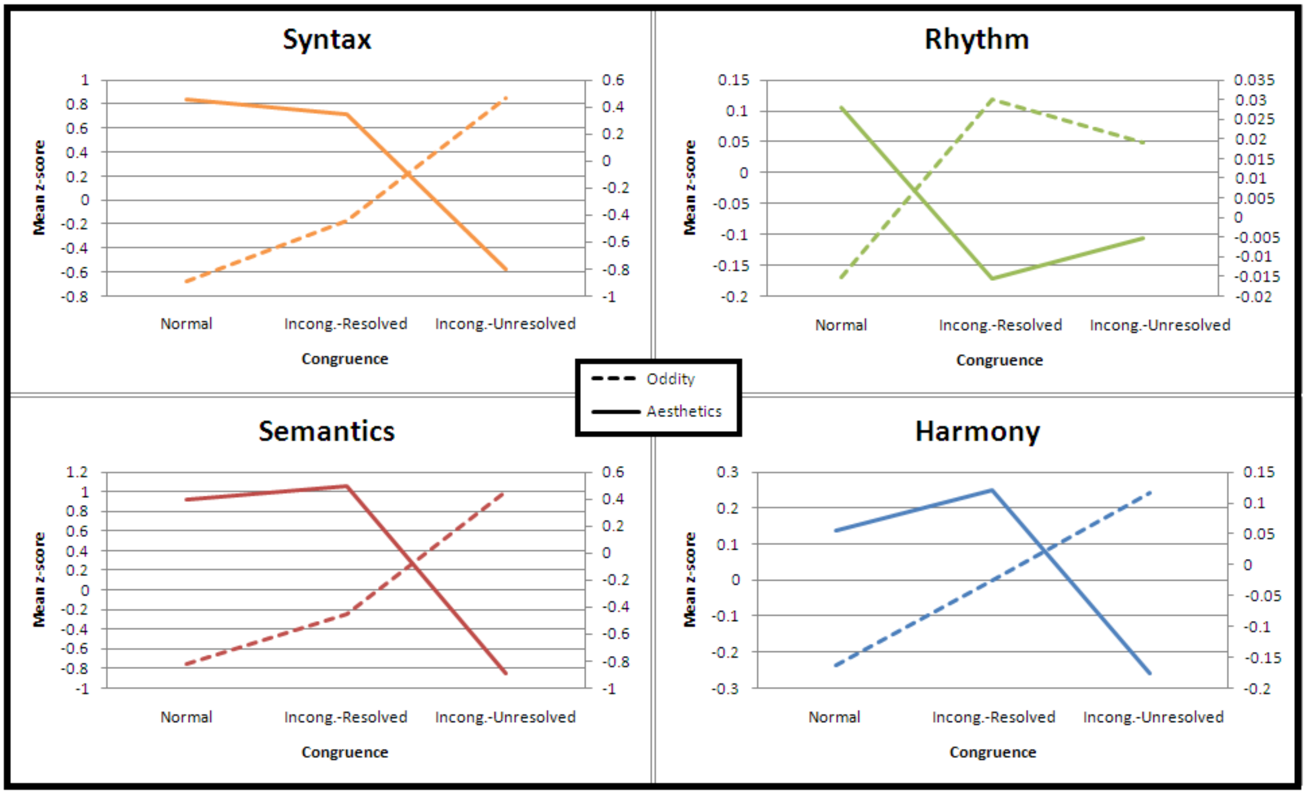

Figure 4.8: Comparison of *oddity* and *aesthetic-affective* rating data trends

Although these correlations could be the result of a genuine effect of the experimental manipulation, this observation raises concerns over whether participants felt they had to report a negative response to a stimulus after reporting that the stimulus was odd. This potential confound was ruled out in Experiment 3 (Chapters 6 to 8) where oddity-related scales were presented in a different experimental block from aesthetic-affective scales.

4.3.5.5 Aesthetic-affective ratings - summary

Despite the slight increase in aesthetic-affective ratings from *normal* to *incongruous-resolved* stimuli in *harmony* and *semantics*, the analyses presented no significant support for the notion that incongruous-resolved stimuli result in a more positive affective response. Participants did, however, show a strong dislike for stimuli in which the end was at odds with the beginning. Once more, similar trends were observed between the *harmony* and the *semantic* stimuli, and the way in which oddity ratings compared to aesthetic-affective ratings showed similarities between *rhythm* and *syntax.* As mentioned at the outset of this section, these scales were intended as a rough estimate of aesthetic-affective experience, and the different patterns observed here suggest that more refined measures are needed to fully capture the different facets of this aspect of music and language aesthetic experiences. Furthermore, the correlation of the aesthetic scale data with the oddity scale data suggest that participants may have been using this scale to denote good form rather than their appreciation of the stimulus, or that the oddity ratings may be influencing the aesthetic ratings.

4.3.6 Summary of results

The data collected in this experiment showed that even in a sample in which most participants had received very little musical training, space-bar presses denoting incongruity detection and average ratings of oddity were significantly higher in stimuli containing an incongruous element than in stimuli containing no incongruity. In addition, the aesthetic-affective ratings showed no significant increase between the ratings of *normal* and *incongruous-resolved* stimuli but a significant decrease in the rating of *incongruous-unresolved* stimuli in all levels of *hrss* bar *rhythm.* Finally, certain patterns of cross-domain similarities emerged, such as a similarity in the effect of the experimental manipulation on the oddity ratings of all four levels of *hrss*, the similarities between the effects of *incongruity-type* in *harmony* and *semantics* stimuli, the parallel trends observed between the aesthetic-affective ratings of *harmony* and *semantics*, and the mirror-image patterns between oddity and aesthetic-affective ratings in *rhythm* and *syntax.*

4.4 Discussion

4.4.1 Oddity ratings

The stimuli used in this experiment were designed to test the assumption that theoretically incongruous elements would be perceived as odd. In *harmony*, *rhythm*, *semantics* and *syntax*, stimuli containing incongruities were rated as significantly more odd than stimuli in which no incongruity was present. These stimuli also received significantly more space-bar presses indicating the detection of an incongruous element. No significant effect of musical training was seen in either the rating data or the space-bar data, supporting the notion that mere exposure to statistical regularities in music and language, rather than the explicit learning of the rules governing good form in these domains of human cognition, was sufficient for participants to be able to determine which stimuli were normal and which were not (Tillmann et al., 2000).

Congruent with expectations, stimuli in which the end was at odds with the beginning, and which therefore could not be integrated into a meaningful whole, were perceived as most odd in *harmony*, *semantics* and *syntax*. The fact that the rhythm stimuli showed no difference between the *incongruous-resolved* and the *incongruous-unresolved* conditions could be explained by some participants hearing both the incongruity and its resolution as odd, since both involve a change of rhythm (as the music transitions into the altered rhythm [the incongruity] and back out of it into the previously established rhythm). This would give participants two points in each *incongruous-resolved rhythm* stimulus at which to notice (and indicate) an incongruity. Furthermore, because rhythmic perception is dependent upon the perception of a stable meter, the meter adopted by participants during the rhythmically ambiguous section of these stimuli is potentially problematic. This explanation would require participants to have switched to a new meter to accomodate the incongruity rather than to have maintained the original beat and thus build up tension between the previously acquire rhythmic framework and the ongoing (framework-incongruous) musical pattern. The space-bar data are consistent with this interpretation, since only in *rhythm* did the *incongruous-resolved* condition obtain a higher average frequency of presses than the *incongrous-unresolved* condition. The inability to determine which meter participants had adopted however calls for caution in this interpretation of the results.

Another explanation for the oddity trend in *rhythm* could be that participants listened out for an incongruity, and made up their mind that the stimulus should be classed as odd once they had encountered it rather than waiting to see whether the incongruity resolved. Once more, since two changes were present (the onset and offset of the incongruous part of the stimulus), participants had two chances to notice an incongruous element in the *incongruous-resolved* condition and only one in the *incongruous-unresolved* condition. The fact that participants were instructed to listen out for incongruous elements could have influenced the strategy adopted for deciding how odd to rate a stimulus. To eliminate this potential confound in oddity ratings, subsequent studies will be less explicit in their reference to what makes the stimuli odd, hopefully encouraging a more global approach to the stimuli.

These accounts suggest an understanding of oddity ratings based upon change detection, rather than upon whether the incrementally processed musical elements can be integrated into a meaningful whole. These interpretations of the *rhythm* data are also at odds with accounts of music psychology influenced by the Gestalt school of thought, which stress the mind's striving for stability and the importance of integrating incoming information into a stable hierarchic structure (Sussman, 2005). However, it could be argued that because rhythmic incongruities occur over time, participants were unable to hold onto the first rhythmic pattern once they had heard the alternative one, rendering the return to the original pattern a further destabilising event rather than a return to the familiar. This would then be a short-term memory issue rather than a musical good-form issue.

The trend in the *harmony* stimuli does not call for this interpretation, since the *incongruous-unresolved* condition did receive a higher mean oddity rating than the *incongrous-resolved* condition. The above account therefore relies on rhythmic incongruities being more salient than harmonic oddities, with participants not being able to pinpoint the elements in the harmonic stimuli which made the stimuli sound odd, or on participants adopting a more global processing strategy in *harmony* than in *rhythm*. Space-bar frequencies in the *incongruous-resolved* condition were indeed slightly higher in *rhythm* than in *harmony* in response to incongruous elements, but this difference was not significant.

However, before settling for any interpretation, it must be noted that this was a first rating study aimed to eliminate any confounding stimulus triplets which did not show good *congruence* condition fitness. The condition fitness in the rhythmic stimuli could have been compromised by the experimental manipulation itself. Adapting a stimulus to fit with an irregular rhythm following the process described in chapter 3 resulted in some cases in a *normal* stimulus which had an irregular rhythm throughout. Further discussion of the strategies underlying oddity judgements must therefore await a study using only the stimuli which showed good condition fitness (see Chapter 7).

Most participants reported not being able to tell the difference between stimuli containing an incongruity and normal stimuli. This was disproved in their ratings, but suggests a lack of awareness of the effect of incongruities on their ratings. Several participants either reported feeling more tense in some stimuli than in others or showed signs of increased tension (such as shoulders lifting or facial expressions suggesting discomfort), prompting the introduction of a "tension felt while listening to the stimulus" scale in subsequent experiments.

4.4.2 Aesthetic-affective ratings

Across all four levels of *hrss*, *incongruous-unresolved* stimuli received the lowest rating: stimuli in which a resolution was not achieved produced negatively valenced aesthetic-affective responses. Contrary to the oddity ratings, this finding is in line with the notion that the lack of a possible integration of elements is unpleasant to participants.

No significant increase in ratings was observed for the *incongruous-resolved* stimuli compared to *normal* stimuli, apart from the *semantics* stimuli when participants were asked to rate their response to the event described in the sentence. Both *harmony* and *semantics* stimuli showed very slight trends in the desired direction, leaving open the possibility of finding support for incongruity-based theories of music and language once the stimulus set has been normed and the questions relating to the aesthetic-affective aspect of participants' responses have been refined.

The deliberate breadth of interpretations of the scales provided in this first experiment led to a number of possible interpretations for the fact that the expected effect was not seen. Considering the mirror image between the oddity ratings and the aesthetic-affective ratings in *rhythm* and *syntax* stimuli, it could be that participants were using the *positive to negative* scale to denote good form rather than their appreciation of the stimuli (i.e. a stimulus which was perceived as odd could not receive a positive aesthetic-affective response). The fact that this mirror image was not as obvious when looking at the data from both scales together in *harmony* or *semantics* would suggest a less salient oddness about these latter two types of stimuli, in line with the suggestion offered for the incongruity trends observed in the *rhythm* stimuli. However, despite the differences between patterns across *hrss*, the correlation between oddity and aesthetic ratings was significant in all levels of *hrss*.

The multiple interpretations of these data were tackled in subsequent chapters of this thesis. The refining of scales through the verbal labelling task (Experiment 2) described in Chapter 5 allows a more precise testing of different facets of participants' responses in Experiment 3. To eliminate any possible contamination of the aesthetic-affective ratings by the participants' memory of how odd they had rated a stimulus, oddity and aesthetic-affective judgements were also separated in Experiment 3 (Chapters 6 to 8).

4.4.3 Comparisons between music and language

Across all levels of *hrss*, the same experimental manipulation, resulting in theoretically equivalent degrees of incongruity, resulted in similar oddity-related behavioural patterns: a higher frequency in space bar presses and higher mean oddity rating scores in stimuli containing theoretical incongruities than in *normal* stimuli.

Other aspects of the analyses showed similarities between *harmony* and *semantics*, and between *rhythm* and *syntax*. The effect of *incongruity-type* showed the same trend in *harmony* and *semantics*. In the aesthetic-affective ratings, both *harmony* and *semantics* stimuli showed a slight increase from *normal* to *incongruous-resolved* while *rhythm* and *syntax* showed a decrease between those two *congruence* conditions. Finally, the relationship between oddity ratings and the aesthetic-affective ratings looked more similar in *syntax* and *rhythm* than in *semantics* and *syntax*, though significant correlations of a similar magnitude were observed across all levels of *hrss*.

4.4.4 Considerations for future studies

The conclusions presented above suggest several areas of improvement for future studies. Though no significant effects of musical training were seen when analysing *years of musical training* as a covariate in the analysis, the removal of this factor from the ANCOVA did affect the outcomes of some analyses. The participant sample did not provide a balanced or controlled means of systematically evaluating the effects of musical training, but this observation lends support to the need to test the effects of musical training on incongruity processing across domains in subsequent experiments. Rather than analysing the effects of musical training in a post hoc manner, strict criteria were applied in the selection of participants for Experiment 3 (Chapters 6 to 8). This allowed the testing of participants only from the two extremes of the musical training spectrum.

The correlations observed in the ratings of oddity and aesthetic-affective responses may have had nothing to do with the juxtaposition of those two scales. However, to eliminate that potential from future studies, ratings relating to different aspects of participants' responses to stimuli were separated in the final experiment. Comparisons between the oddity scale and scales denoting different facets of the aesthetic-affective response are compared in chapter 7 to see whether these trends were the result of the juxtaposition or whether they reflected a true effect of the experimental manipulation.

Because of the systematic approach adopted in the creation of the stimulus set, some of the stimuli designed to sound *normal* may have failed to sound completely normal. In creating the stimuli, the intention was always to refine the stimulus set based on participants' responses in Experiment 1. Subsequent experiments in this thesis used the refined stimulus set, from which stimuli showing poor *congruence* condition fitness were removed using the procedure described in Chapter 5. Finally, the vagueness of the aesthetic-affective rating scales led to difficulties in interpreting the data collected in this first experiment. These scales were only intended to serve as a rough measure of participants' aesthetic-affective responses, with the intention of developing a more precise set of scales. This development, described in Chapter 5, used a verbal labelling task. These two key steps in this project ensure a bottom-up approach to refining both the research question and the means of addressing this aspect of music and language cognition.

Chapter 5

Norms and labels: refining the stimulus set and the behavioural scales for the final experiment

5.1 Introduction

Although Experiment 1 showed overall good *congruence* condition fitness, it had been decided from the start that the data from the experiment reported in chapter 4 would be used to refine the stimulus set. The norming procedure described below indicates how this process was carried out in a systematic manner to eliminate any stimulus triplets which did not show good *congruence* condition fitness in a by-items analysis. A second aspect of this research which required refinement before the final study was the measures used to capture participants' behavioural responses to the stimuli. These were developed in a bottom-up approach following a verbal labelling task. This chapter details both these refining processes.

5.2 Norming procedure

In the interest of selecting only stimuli which showed good *congruence* condition fitness, the stimulus set described in Chapter 3 was subjected to a strict norming procedure, using the data from the rating experiment reported in Chapter 4. Overall, this method of developing stimuli combines methods used in music psychology and in psycholinguistics. In the former, the stimuli are considered more or less incongruous according to the extent to which they adhere to the rules and probabilities of Western tonal harmony (Patel et al., 1998; Koelsch et al., 2001). In the latter, stimuli are often piloted in the form of sentence completion tasks or grammaticality judgement tasks (e.g. Van Berkum et al., 2005). Completion tasks are used to ascertain which sentence continuation is expected by participants. Grammaticality judgement tasks are used to select sentences which are perceived as more or less odd, or which are more or less difficult to understand. Completion tasks are notoriously difficult to carry out in music (Schmuckler, 1989), and correctness judgement tasks result in similar outcomes as completion tasks (Schmuckler, 1989). The data from Experiment 1, in which participants were asked to judge how "odd" the stimuli were, was therefore used to refine the stimulus set for future experiments. These data were used in a series of evaluations aimed at assessing how "odd" each stimulus was perceived to be.

Only the 96 triplets in which the three stimuli in each level of *congruence* showed the best fit were retained for the ERP experiments in this series of studies. This good condition fit was defined as follows: for the *normal* stimuli the behavioural data should indicate no perception of any incongruous element and a low rating of oddity; for the *incongruous-unresolved* stimuli, the data should show a high rating of oddity and an indication that participants had perceived the manipulated incongruity as odd; for the *incongruous-resolved* condition, the data should show an oddity rating higher than the *normal* stimuli but lower than the *incongruous-unresolved* stimuli, and an indication that the experimentally manipulated incongruity was perceived as unusual.

In summary, good *congruence* condition fitness was demonstrated by the *normal* stimulus being perceived as normal, the *incongruous-resolved* stimulus being perceived as unusual but not incomprehensible and the *incongruous-unresolved* stimulus being perceived as very odd, within each triplet, as a result of the manipulated incongruity, not because of a different part of the stimulus. In order to eliminate the possibility of stimuli being selected for future studies on the basis of participants' positive or negative affective responses (which lie at the centre of the main research question of the thesis), the selection process followed a systematic approach aimed at assessing only how odd stimuli were perceived to be.

5.2.1 Successive evaluations

Stimuli were selected according to the scores they received in evaluations, described below, designed to test their goodness of fit with the three levels of *congruence*: *normal*, *incongruous-resolved* and *incongruous-unresolved*. Failing an evaluation resulted in a -1 score, passing an evaluation resulted in a $+1$ score. This ensured that passing one evaluation and failing on another resulted in a score of 0. In addition, a *boost* category was developed for each evaluation, giving a score of $+2$ to stimuli which showed the ideal data pattern of an evaluation. This ensured that stimuli which were the best according to each evaluation received the highest total score.

The evaluations were designed to assess the validity of the different stimuli within a triplet separately. Evaluations 1, 2 and 3 assessed the validity of the *normal* stimuli, evaluations 4 and 5, the *incongruous-resolved* stimuli and evaluations 6 and 7 the *incongruous-unresolved* stimuli. It was therefore possible to compute subtotal scores for each stimulus within a triplet. This allowed the selection of stimulus triplets which performed well across all levels of *congruence*, rather than allowing a stimulus which achieved a very high score in two *congruence* conditions and a very low score in a third condition (obtaining an average to high overall score) to compromise the outcome of the norming study.

5.2.1.1 Evaluation 1: Do any of the stimuli in the *normal* condition show a cluster of space-bar presses?

Rationale If more than one participant presses the space bar within a period of 1000ms on the same stimulus which is designed not to contain any incongruities, the stimulus is considered to contain an element which can be perceived as "odd" and should be rejected as part of the *normal* condition. The decision to maintain a minimum of two presses was made to avoid excluding stimuli on the basis of a random error presses.

Boost Stimuli which show no space-bar presses at all.

Fail Stimuli which show two or more space-bar presses in a time window of less than 1000ms.

5.2.1.2 Evaluation 2: Do any of the stimuli in the *normal* condition show false positives for more than 50% of participants?

Rationale If more than 50% of participants pressed the space-bar during the hearing of a stimulus, this indicates that less than 50% of participants consider that this stimulus sounds normal, even if they disagree about the point at which the abnormality occurs. This would show a bad condition fit, since the *normal* stimulus is not reliably perceived as normal.

Boost Stimuli with less than 20% space-bar pressed trials (10% for language stimuli, adjusted because of low overall space bar press frequencies in the language stimuli).

Fail Stimuli for which 50% or more participants pressed the space bar.

5.2.1.3 Evaluations 3 (*normal*) and 4 (*incongruous-resolved*): Is the rating trend (normal - slightly unusual - very odd) in the expected direction?

Rationale Although some participants reported that they found the space-bar press task difficult, the ratings which they provided allowed them to indicate that a stimulus sounded odd even if they omitted to press the space bar. The optimal ratings for these stimuli should show the *normal* stimulus rated less odd than the *incongruous-resolved* stimulus, which in turn should be rated less odd than the *incongruous-unresolved* stimulus.

Boost Stimuli with the trend in the expected direction (*normal* $<$ other two conditions for *normal* boost, *normal* $<$ *incongruous-resolved* $<$ *incongruous-unresolved* for *incongruous-resolved* boost).

Fail for N scoring Stimuli in which the *normal* version is rated more odd than either of the other two versions of the same stimulus.

Fail for O scoring Stimuli in which the *incongruous-resolved* version is rated less odd than the *normal* version.

5.2.1.4 Evaluation 5: In the *incongruous-unresolved* condition, do participants press the space-bar only once the incongruity has occurred?

Rationale If the space bar is not pressed during the hearing of the stimulus, it could be that participants do not consider the stimulus odd. If the space bar is pressed before the occurrence of the incongruity in the stimulus, participants are classing the stimulus as odd for reasons other than the experimental manipulation. In both cases, the stimulus must be rejected.

Boost Stimuli in which the space bar is pressed in over 50% of trials after the incongruity has occurred.

Fail Stimuli in which the space-bar is pressed by less than 20% participants after the occurrence of the incongruity and in which more than 20% participants press the space bar before the incongruity (10% for language stimuli, mirroring the adaptation of evaluation 2).

5.2.1.5 Evaluation 6: Is the space-bar pressed at an appropriate time in the *incongruous-resolved* condition?

Rationale In the *incongruous-resolved* version of each stimulus, the portion of the stimulus designed to sound unusual is between the onset of the incongruity and its resolution. Considering that the resolution forms part of the element of interest, but that beyond the resolution the stimulus should be considered normal, good condition fitness would consist in participants pressing the space bar only as a result of this portion of the stimulus: between the onset of the incongruity and a second after its resolution, to allow a reasonable reaction time.

Boost Stimuli in which over 75% of space-bar presses are within an "incongruity to resolution + 1000ms" time window.

Fail Stimuli in which over 50% of space-bar presses are outside that range.

5.2.1.6 Evaluation 7: Is the *incongruous-unresolved* version of the stimulus rated as more odd than the other two versions of the stimulus?

Rationale The *incongruous-unresolved* versions of the stimuli are designed to be impossible to compute into a meaningful unit. Their ratings on the oddity scale should therefore be high.

Boost Stimuli in which the *incongruous-unresolved* stimulus is rated more odd than the other two stimuli within its triplet

Fail Stimuli in which the *incongruous-unresolved* version is rated less odd than the *normal* version

5.2.2 The selection process

A number of observations can be made at this point in the selection process. Firstly, scores were generally higher in the language stimuli than in the music stimuli. This can be attributed to a greater degree of expertise in noticing incongruities in language as compared to music, considering that all participants were native English speakers but that most participants had less than four years of musical training. Secondly, no single evaluation carried the weight of the rejection or acceptance decision. The fail scores were highest in different evaluations within all four levels of *hrss* for different stimulus triplets, and stimuli which failed on only one evaluation did not all fail on the same evaluation.

Consequently, the evaluations of the stimuli were all given equal weighting in the selection process as no evaluation stood out as most representative of good condition fitness. Stimuli were selected in an incremental way, with the aim of obtaining a final stimulus set which contains only stimuli which achieve a high evaluation score in all conditions. The following systematic procedure was adopted to ensure the selection of the best 24 stimulus triplets in each level of *hrss*s (*harmony, rhythm, syntax, semantics*).

1. Reject stimulus triplets with an overall fail score.
2. From the remaining triplets, reject those which fail on any given *congruence* condition.
    (a) If this brings the number of triplets below 24, keep the stimuli with the highest overall score, preferring those with similar subtotal scores to those with a strong fail in one condition.
    (b) If the total number of triplets is still above 24, proceed to next step.
3. From the remaining triplets, keep the stimuli with the highest overall scores, preferring those with a highest total score on the evaluations associated with the *normal* condition in the case of a tie.

5.2.3 Outcome

The scores given to the stimuli enabled the selection of only stimuli which achieved an overall pass. In the semantics stimuli, no stimulus triplets obtained an overall score of less than 0. In the syntax stimuli only one triplet was rated as -1. In the harmony stimuli, 10 triplets failed to meet the criteria. In these stimuli, the fails were equally distributed across the incongruity types ($a = 3, b = 4, c = 3$), suggesting that no incongruity type systematically resulted in unreliable stimulus triplets. In the rhythm stimuli, 8 triplets achieved an overall fail, with a predominance of the fails being in c type stimuli ($a = 2, b = 1, c = 5$). This can be explained by the fact that a c type incongruity in rhythm resulted in an unusual number of beats in a bar, requiring the original meter to be temporarily suspended and therefore making the rhythmic resolution problematic. Furthermore, when the piece was adapted to make the incongruity fit within its context, all the bars in the piece contained an unusual number of beats, making the whole piece sound odd in the *normal* condition, and disrupting the rest of the ratings. This was anticipated upon stimulus creation and only emphasises the fact that incongruity detection is not merely down to matching an element with its context but relies to a certain degree on matching an element to what one knows to be normal through past experience.

Post-selection, the split of stimulus triplets between the three types of divergence from the norm described in section 1 was as follows: Harmony ($a = 9, b = 7, c = 8$), Rhythm ($a = 9, b = 10, c = 5$), Semantics ($a = 8, b = 7, c = 9$), Syntax ($a = 8, b = 9, c = 7$). The end result was therefore a fairly even representation of all three divergence types in all stimulus types apart from the rhythmic stimuli in which fewer c-type stimuli were retained. The explanation given above for the greater fail rate of this type of stimuli serves to explain this outcome, since overall-fail stimuli were automatically rejected from the final stimulus triplet. The stimuli within that type which did not fail were still likely to do less well in the evaluations than other rhythmic stimulus triplets, due to the unusual number of beats in a bar in the *normal* condition. Though it could be argued that it would have been better to start with a set of stimuli in which the *normal* condition in *rhythm* always contained a regular beat, this systematic approach in the creation of stimuli allowed the investigation of whether oddity was rated as a function of novelty or template-matching.

Under the first account of oddity ratings, a piece in which each bar was in 5/4, or even in which the sequence consisted of a systematic alternation of bars of 5/8 and 6/8 (leading to an overall consistent beat of 11/8) would not be perceived as odd. Under the second account, pieces with such an unusual rhythmic structure would be perceived as odd throughout, since the participants would be unable to ground the incoming stimuli onto a rhythmic structure to which they have grown accustomed through exposure to the music of their culture. Since stimuli in which the overall beat was unusual (e.g. the whole piece was in 5/4) were rated as odd even when there was no change to the rhythmic structure of the piece for the duration of the piece, the analysis carried out here points towards a template-based perception of oddity.

No significant effect for musical key was found for the rating of harmonic and rhythmic stimuli ($p > 0.05$). In the interest of maintaining an equal representation of all musical keys in future studies, and to limit repetition of keys within the stimulus triplet, four stimuli were transposed following the selection process. These transpositions were minimal (only a small number of semitones) so as not to disrupt the voicing of the stimuli (one of the rules of classical harmony stipulating no thirds below the F# below middle C, a large transposition downwards could for instance have caused the stimulus to violate a rule with which it was compliant pre-selection). The outcome of these transpositions was a post-selection stimulus triplet in which all 12 keys of western tonal harmony were equally represented.

The music stimuli retained post-selection did not have an equal representation of the three tempi used to create the stimuli (rhythm, *70* = 9, *100* = 8, *130* = 7; harmony, *70* = 5, *100* = 8, *130* = 11). The tempo affected the condition fit more in the harmonic stimuli than in the rhythmic stimuli, with a greater fit for faster music in *harmony*.

An analysis pre-selection revealed a significant main effect for tempo on the oddity ratings in both the harmony ($p < 0.001$) and the rhythm stimuli ($p < 0.02$) as well as a significant *condition* x *tempo* interaction effect ($p < 0.001$). This was initially deemed problematic, since any tempo changes carried out on the stimuli post-selection to ensure an equal distribution of all three tempi might invalidate the selection process. A post-selection analysis however revealed that the main effect for tempo had been rendered non-significant by the selection process ($p = 0.81$ for harmony, $p = 0.44$ for rhythm). The *tempo* x *condition* interaction effect was also rendered non significant ($p = 0.59$). Since no significant main effect for tempo was found in the post-selection stimuli, the decision was made to leave the tempi as they were, rather than risking invalidating the selection process, resulting in a set of stimuli in which the three tempi, though unequally represented, had no significant effect on the ratings. For obvious reasons, these considerations of key and tempo were irrelevant to the language stimuli. The lists of stimuli in all levels of *hrss* with their selection scores and keys pre- and post- selection where relevant are presented in Appendix B.1.

5.2.4 Discussion

Using total scores as a measure of condition fitness, this procedure allowed the selection of the fittest 24 triplets of stimuli derived from the same incongruity within each level of *hrss*. Total scores on the test criteria were generally higher in the language stimuli than in the music stimuli. This can be attributed to a greater degree of expertise in noticing incongruities in language as compared to music, since all participants were native English speakers and therefore had much greater expertise in hearing and manipulating language than music. No single evaluation carried the weight of the rejection or acceptance decision across all stimulus triplets, and the congruency of passes and fails within any given triplet supported the case for successive evaluations.

A re-analysis of the data collected in Experiment 1 (chapter 4) was conducted using only the stimuli selected via this procedure. This analysis revealed stronger effects for *congruence* in the music stimuli than were found when the entire pool of stimuli were used: for *harmony* and *rhythm* respectively, pre selection, $\eta^2 = 0.45$ and $\eta^2 = 0.19$ and post selection $\eta^2 = 0.52$ and $\eta^2 = 0.21$. The strength of the main effects for condition in the language stimuli varied by less than $\eta^2 = 0.1$. Two potential artefacts however compromise the validity of the post-norming reanalysis of ratings. Firstly, it could be argued that the selection of stimulus sets which are in line with the desired trends would automatically lead to improved trends. This first artefact alone calls for a re-test reliability study. Secondly, participants' ratings may be affected by the pool of stimuli in which the spectrum of the oddity scale was distributed. If, for instance, the stimulus pool contained mostly odd-sounding stimuli, the least odd of the odd-sounding stimuli would receive a low oddity rating (indicating that the stimuli were perceived as near-normal). If, on the other hand the stimulus pool contained mostly normal-sounding stimuli, these same marginally odd stimuli would receive a high oddity rating. Consequently, a validation of the condition-fitness displayed in this stimulus set requires new data. These data are collected in the final experiment of this thesis, in which participants heard only the stimuli selected in this norming procedure and rated them on the same oddity scale as used in the initial rating study (see Chapter 7).

5.3 Developing a set of scales to capture participants' aesthetic-affective responses

The initial rating study described in chapter 4 used a *negative to positive* rating scale as a rough indicator of participants' aesthetic-affective ratings following each stimulus. The analyses of the data in chapter 4 showed a different relationship between the oddity ratings and the aesthetic-affective ratings for different levels of *hrss*, as well as a different effect of *congruence* across levels of *hrss*. Though this could be a genuine effect of the experimental manipulation in *harmony* and *semantics* in comparison to *rhythm* and *syntax*, the nature of the scale made several interpretations possible. The variability in the data, though partly due the deliberately vague nature of the scales, allowed the possibility that participants' responses encompassed a wide range of responses including aesthetic appraisals, grammaticality judgements and affective responses.

This outcome confirmed the need for more refined scales for further studies. Means of measuring aesthetic-affective responses vary hugely across published studies. Many studies in the area of music and language cognition have focussed on tension and relaxation (Steinbeis et al., 2006; Vuust et al., 2006). Others using ratings for emotions have assessed pre- and post-stimulus mood (Gfeller et al., 1991; Stratton & Zalanowski, 1994; Ali & Peynircioglu, 2006). Others still have studied emotional responses though a semantic priming paradigm (Steinbeis & Koelsch, 2008b; Koelsch, 2008). None of these measures fully captured the nature of the aesthetic affective responses of interest in this project. Instead, they captured tension, mood induction and, according to the authors, semantic priming of emotion-related words, but none focussed on the immediate affective change in response to a specific musical or linguistic element. Furthermore, the cross-domain validity of these scales is questionable. Since this project aims to compare the role of incongruities across music and language, it was essential that the scales reflect words used by participants to describe their responses to incongruities encountered in both domains.

In order to create scales which would capture the types of affective and cognitive responses elicited by the manipulation of music and language incongruities, a bottom-up approach was adopted. A verbal labelling experiment was carried out in which participants were asked to record a few words which best captured their response to each of the stimuli they heard. Data were analysed to ascertain the prevailing dimensions in descriptions of participants' responses to both music and language stimuli. Though it could be argued that this resulted in a loss of richness in the data, the necessity of creating scales which could be used for ratings across both domains in subsequent experiments justified this decision. Furthermore, the studies reviewed in the last section of chapter 2 suggested similar types of responses elicited by the processing of incongruities across domains. Those were the responses which were central to this research.

This bottom-up approach was adopted in an aim to increase the validity of the measures used to capture aesthetic-affective responses — not only were the labels for future scales provided by the participants, and thus understandable by future participants, this also allowed a much more refined investigation of the different facets of responses to music and language.

5.3.1 Experiment 2

Eighteen participants took part in this verbal labelling experiment during an open day held by the Institute of Psychological Sciences at the University of Leeds (10 female, mean age = 61.06, mean years of music training before adulthood = 3.61). The average age of the participants is explained by the demographic of the people able to attend an open day on a day when most people of working age would be at work. The discrepancy of ages between this set of participants and the participants taking part in other experiments in this thesis was a slight cause for concern. However, participants in the final experiment (Experiment 3) reported a good understanding of the labels generated by the participants in Experiment 2.

Participants listened to 48 stimuli in total, consisting of 24 stimuli from each level of *hrss* (*harmony* and *syntax* or *rhythm* and *semantics*). These 24 stimuli were one stimulus from each triplet selected in the norming procedure described in the previous section. Participants therefore heard an equal number of stimuli in each *congruence* condition from two levels of *hrss* selected through the norming procedure. The experiment used a 2 (*hrss*) x 3 (*congruence*) repeated measures design. The order in which participants completed the music or language task was counterbalanced between the two sessions. The stimuli were presented via headphones using E-Prime. All instructions were presented on the screen and participants' responses were recorded in E-Prime.

Participants were tested in groups in a computer laboratory in the Institute of

Psychological Sciences, University of Leeds, in which they were instructed to sit in front of a computer and place a pair of headphones over their ears. All subsequent instructions appeared on the screen. Participants were told that the experiment was a memory task and that they were to listen to 12 stimuli in the learning phase, providing their "gut reaction" to each stimulus by typing a few words into a box which appeared on the screen after the stimulus had finished. The decision to disguise the experiment as a memory task was made to encourage participants to listen attentively to each stimulus. They were instructed to repeat the procedure in the recognition phase: as well as answering a yes or no to the question "did you hear this stimulus in the learning phase?", they were provided with the same empty text box on the screen in which to type their response. Six of the learning-phase stimuli were repeated in the recognition phase and 12 were new. This procedure allowed participants to listen to, and provide responses for, all 24 stimuli of two stimulus types.

5.3.2 Analysis

The first stage of the analysis consisted in identifying labels which had been used to describe responses to both music and language. Table 5.1 presents this set of labels and the *hrss* condition in which they were used by participants to describe their responses to the stimuli.

In the second stage, two independent researchers grouped the labels into categories based on semantic closeness, adopting the following systematic procedure:

1. from the above list, indicate which labels seem to fall into the same semantic category (labels can be used as many times as necessary)

2. provide a description or a label of the constructed categories

The word "seem" in the first instruction was included because of the ambiguity of some of the labels. For instance, if a participant responded "peaceful", it is difficult to say for sure whether this label should be applied to the stimulus or to the

Table 5.1: Labels used by participants to describe their responses to both music and language

Label	Levels of *hrss* in which it was used
(un) excited / exciting	*harmony, semantics*
(un) interested / -ing	*harmony, rhythm, semantics, syntax*
annoyed / annoying	*harmony, rhythm, syntax*
anxious	*harmony, semantics*
bored	*harmony, syntax*
boring	*harmony, rhythm, syntax*
calm / calming	*harmony, rhythm, semantics, syntax*
comical	*rhythm, syntax*
complex	*rhythm, syntax*
confused	*harmony, rhythm, semantics, syntax*
don't know	*rhythm, syntax*
dull	*harmony, syntax*
easy	*rhythm, syntax*
good	*harmony, rhythm, semantics, syntax*
happy	*harmony, rhythm, semantics, syntax*
indifferent	*harmony, semantics, syntax*
irritating	*harmony, semantics*
like / dislike	*harmony, rhythm, syntax*
nice	*harmony, semantics*
none	*harmony, semantics, syntax*
not sure	*rhythm, syntax*
nothing	*harmony, semantics, syntax*
ok	*harmony, rhythm, semantics, syntax*
peaceful	*harmony, rhythm, semantics, syntax*
pleasant / -ing	*harmony, rhythm, semantics*
sad	*harmony, rhythm, semantics, syntax*
simple	*rhythm, syntax*
thoughtful	*harmony, rhythm, syntax*
tired	*rhythm, syntax*
unsure	*rhythm, syntax*

listener; likewise with words such as "good", "ok" or "dull". The second instruction aimed to arrive at phrases or category-labels which could summarise the labels grouped together through the first instruction. Categories which contained words which were already all included in another existing category were collapsed into the existing category. For instance, coder B created the category "physical alertness" which contained the words "tired" and "dull". These words were already included in the category "arousal". To retain the nuances drawn out by different groupings within bigger categories, the category-labels were added to each-other. In the case outlined here, the category “arousal” became “arousal / physical alertness”.

Table 5.2: Summary table demonstrating a strong overlap between coders A and B in the assignment of labels to emerging categories

Dimension	Tension		Stimulation			Confusion/Interest				Aesthetic pleasantness				
Coder	A	B	A	B	B	A	A	B	B	A	A	B	B	B
(un)excited / exciting	✓	✓	✓	✓	✓							✓		
(un)interested / -ing			✓			✓		✓	✓					
annoyed / annoying	✓	✓									✓	✓	✓	
anxious	✓	✓												
bored / boring		✓	✓		✓	✓						✓		
calm	✓	✓	✓	✓							✓			
comical	✓											✓		
complex						✓		✓						
confused		✓				✓	✓		✓					
don't know	✓						✓		✓					
dull			✓	✓	✓	✓	✓		✓			✓		✓
easy	✓	✓		✓		✓		✓						
good	✓									✓		✓	✓	✓
happy	✓												✓	✓
indifferent			✓		✓		✓		✓	✓	✓		✓	
irritating	✓											✓		
like / dislike										✓	✓		✓	
nice										✓	✓	✓	✓	✓
none			✓		✓		✓							
not sure	✓	✓					✓		✓					
nothing		✓	✓		✓		✓							
ok						✓				✓		✓	✓	✓
peaceful	✓	✓		✓								✓		
pleasant / -ing										✓	✓	✓	✓	✓
sad	✓	✓										✓		✓
simple			✓			✓		✓						
thoughtful			✓			✓		✓	✓					
tired			✓	✓										✓
unsure	✓	✓	✓				✓		✓					

Table 5.2 presents the groupings performed by both coders, placing side by side groups which showed a strong degree of overlap in the labels included. The shaded rows are those representing labels used by both coders within these corresponding categories, demonstrating a strong degree of overlap between the two coders in this task when all the categories are juxtaposed.

5.3.3 Outcome

From the process described above, four clear dimensions emerged, dictated by the words used by both coders within each overarching category. Tables 5.3, 5.4 and 5.5 are a more detailed view of Table 5.2, listing the labels within each group formed by the two coders as well as the names of the sub-categories constructed by the coders.

Table 5.3: Detailed table of classification - Tension and stimulation

	Tense / relaxed	**Tension**	**Stimulating**	**Arousal Physical alertness**	**(Un)exciting Interested(ness)**
	Coder A	Coder B	Coder A	Coder B	Coder B
(un)excited / exciting	(un)excited / exciting	(un)excited / exciting	(un) excited / exciting	(un) excited / exciting	(un)exciting / ed
(un)interested / -ing			(un)interested / -ing		
annoyed / annoying	annoying / annoyed	annoyed			
anxious	anxious	anxious			
bored / boring		bored	bored / boring		boring
calm	calm / calming	calm	calm	calm	
comical	comical				
complex					
confused		confused			
don't know	don't know				
dull			dull	dull	dull
easy	easy	easy		easy	
good	good				
happy	happy				
indifferent			indifferent		indifferent
irritating	irritating				
like / dislike					
nice					
none			none		none
not sure	not sure	not sure			
nothing		nothing	nothing		nothing
ok					
peaceful	peaceful	peaceful		peaceful	
pleasant / -ing					
sad	sad	sad			
simple			simple		
thoughtful			thoughtful		
tired			tired	tired	
unsure	unsure	unsure	unsure		

The first label group was one which both coders described as "tension" or "tense / relaxed". Table 5.3 shows a strong degree of overlap between the words classified as tension-related by both coders. Words mentioned by both researchers only in this dimension were "anxious", "peaceful", and "sad". Emotions such as "happy" and "sad" were closer to mood changes investigated in previous research than to the aesthetic affective experiences at the heart of this research project. The word "tense" was therefore retained as it provided a scale in which the extremities would be able to account for feelings of angst and peacefulness, as well as a means of replicating the methods used by other researchers in the field.

A second dimension which emerged from this analysis was one described by the coders as "stimulating", "arousal / physical alertness" and "(un)exciting / interested(ness)". Once more, the words used by both coders in this dimension alone (exciting, boring, calm, dull, indifferent, none, nothing, tired) were compared to words used by both coders in other dimensions. Unique to this dimension were the words "bored", "none", "nothing" and "tired". The word "stimulating" was therefore used to capture this dimension, allowing an interpretation in terms of both intellectual stimulation (which would alleviate boredom) or physical stimulation (which would either increase a participant's alertness or make them tired through too much stimulation).

The third dimension showed the greatest number of unique agreements between the coders (interesting, complex, confused, don't know, simple, thoughtful). It seemed that these agreements showed two aspects, captured by the category labels provided by coder B: one to do with the complexity of the stimulus ("cognitive judgements [object-focussed] / complexity") and one to do with the resulting effect on the participant's cognitive state ("cognitive judgements [self-focussed]"). The category labels provided by coder A highlighted the notion of cognitive processing difficulty ("mentally challenging") and indecision. The words which showed the highest prevalence in this dimension were "interesting", "dull", "confused", and "thoughtful". Considering that these also epitomised the two aspects of this dimension, it was decided to create two scales from this dimension. These were based on how interesting the stimulus sounded and how confused the participant was as a result of hearing it. The word "dull" appeared in a number of categories in the coding exercise, partly because of its ambiguity in terms of whether it described an emotional state (dull, numb, down) or the stimulus (dull, boring, uninteresting). The opposite end of the scale was therefore changed to "bland".

Table 5.4: Detailed table of classification - Cognitive judgements

	Cognitively challenging Coder A	**Indifference Undecided** Coder A	**Cognitive judgement object-focussed Complexity** Coder B	**Cognitive judgements (self-focussed)** Coder B
(un)excited / exciting				
(un)interested / -ing	(un)interesting / -ed		(un)interesting	(un)interested
annoyed / annoying				
anxious				
bored / boring	bored / boring			
calm				
comical				
complex	complex		complex	
confused	confused	confused		confused
don't know		don't know		don't know
dull	dull	dull		dull
easy	easy		easy	
good				
happy				
indifferent		indifferent		indifferent
irritating				
like / dislike				
nice				
none		none		
not sure		not sure		not sure
nothing		nothing		
ok	ok			
peaceful				
pleasant / -ing				
sad				
simple	simple		simple	
thoughtful	thoughtful		thoughtful	thoughtful
tired				
unsure		unsure		unsure

The word "thoughtful" could be understood as full of thought, thought-provoking or considerate. Consequently, the word "confused" was kept as an anchor for the second scale of this dimension.

The final emerging dimension captured notions of judgements concerning the stimuli and pleasantness. Words which participants agreed upon within this dimension alone were "good", "like", "nice", "ok", and "pleasing / pleasant", with "pleasing / pleasant" being the word with the highest frequency within this dimension. These labels revolve around aesthetic judgments (good, ok, nice) and pleasantness (like, nice). The phrase used to capture this dimension was therefore "aesthetic pleasantness".

Presented in order of the dimensions described above, five scales were used to capture

Table 5.5: Detailed table of classification - Aesthetics and affect

	Aesthetic evaluation	Pleasantness	Distant judgement object focussed (Un)amusingness	(Dis)pleasingness (Un)pleasantness Appealingness	(Un)happiness
	Coder A	Coder A	Coder B	Coder B	Coder B
(un)excited / exciting			(un)excited / exciting		
(un)interested / -ing					
annoyed / annoying		annoyed / annoying	annoyed / annoying	annoyed / annoying	
anxious					
bored / boring			boring		
calm		calm			
comical			comical		
complex					
confused					
don't know					
dull			dull		dull
easy					
good	good		good	good	good
happy				happy	happy
indifferent	indifferent	indifferent		indifferent	
irritating			irritating		
like / dislike	like / dislike	like / dislike		like / dislike	
nice	nice	nice	nice	nice	nice
none					
not sure					
nothing					
ok	ok		ok	ok	ok
peaceful			peaceful		
pleasant / -ing	pleasant / -ing	pleasant / -ing	pleasing / pleasant	pleasant / pleasing	pleasant
sad			sad		sad
simple					
thoughtful					
tired					tired
unsure					

the different facets of participants' responses in the final experiment of this thesis.

- How do you feel now, having listened to the whole stimulus? (very relaxed — very tense)
- How stimulating did you find the stimulus? (not at all / intensely)
- In your opinion, the stimulus was: (very bland — very interesting)
- How confused or perplexed do you feel, having heard the whole stimulus? (not at all — very)
- How aesthetically pleasing was the stimulus as a whole? (not at all — very)

To allow a validation of the stimulus norming procedure, and to ensure that the stimuli showed good condition fitness in the oddity ratings, the oddity scale used in the initial rating study was also used in the final study. Finally, to reflect participants' feedback and the observation of the behaviour of participants taking part in the initial rating study (see the discussion section in chapter 4), a scale concerning tension felt while listening to the stimulus was added. The two following scales were therefore added to those developed via the verbal labelling task described in this chapter:

- How odd was the stimulus? (completely normal — very odd)
- How tense did you feel while listening to the stimulus? (not at all / very)

These scales were used in the final study of this thesis to allow an analysis of behavioural data alongside ERP data, in the pursuit of the role of incongruities in eliciting aesthetic-affective responses to music and language.

5.3.4 Discussion

The coding and grouping of verbal labels used to describe participants' responses to both music and language led to four clear dimensions from which scales could be constructed. It could be argued that discarding labels which were only applied to music or language may have resulted in a skewed representation of the responses elicited by incongruities within each domain. However, since the present project is concerned with the role of incongruities in aesthetic-affective responses across both domains, only the labels which could be applied to both domains are of interest here. Furthermore, these labels seemed to capture the responses described in theories of incongruity-based music and language aesthetics: tension (Steinbeis et al., 2006; Vuust et al., 2006), arousal (Steinbeis & Koelsch, 2008b; Koelsch, 2008), and aesthetic pleasantness (Flores-Gutierrez et al., 2007; Koelsch et al., 2008). Because of concern of oddity judgements affecting subsequent aesthetic-related judgements expressed in response to the data presented in

Chapter 4, scales denoting these two aspects of participants's responses to the stimuli were presented in separate experimental blocks in the final experiment. These data enable the study of the relationship between the different facets of responses to aesthetically incongruous stimuli (Chapter 7) as well as between these behavioural measures and ERP effects (Chapter 8). The final experiment was designed not only to see the impact of different contributors to aesthetic experiences across domains, but also with the hope of showing which aspects of participants' conscious subjective responses bear the closest relationship to the ERP patterns collected for the same stimuli, to help regroup and refine the currently multiplying interpretations of ERP effects.

Chapter 6

Sounding the sparks: capturing the effect of incongruities through the collection of behavioural data, ERP data, and the investigation of the relationship between the two

6.1 Introduction

This chapter presents the rationale for, and the methods used in, Experiment 3. This extensive experiment, which collected both EEG data and behavioural data was the culmination of the research carried out up to this point in the thesis. Its aims were to discover what specific elements of participants' responses, behavioural and neurophysiological, underpinned aesthetic-affective responses to incongruities in music and language, and to compare these contributors across the four components of music and language. The first aim enabled a contribution towards the call for more systematic investigation of incongruity-based models of aesthetics within each domain, while the second allowed the testing of whether the same model can be applied across both domains. Secondary aims included an exploration of the interpretation of ERP effects, by examining ERP effects and behavioural effects side by side, and an exploration of the effects of musical training on both music and language processing.

6.2 Rationale

6.2.1 Questions arising from the research carried out thus far

The systematic review of the literature carried out at the onset of this thesis led to several research questions. The first was the question of whether similar patterns would be observed in response to aesthetically-purposed incongruities in music and language, both in behavioural outcomes and in neurophysiological outcomes. The impetus for this research question was provided by the similarities observed in behavioural, cognitive, affective and neural patterns in both domains. A second question was whether the notion of harmonic closure could account for the differences between studies of harmonic incongruity processing which report either a late positivity (e.g. Patel et al., 1998) or a late negativity (e.g. Koelsch et al., 2003). A third question, tied to the previous observation, concerned the interpretation of ERP components, and the question of whether the association of greater ERP effect amplitude with a perception of oddity in stimuli would be empirically supported.

The unexpected pattern in the *rhythm* stimuli in the rating study, led to a fourth question, namely whether both the onset of the rhythmic incongruity and the return to the original pattern were perceived as incongruous. Though the space-bar data suggested this could be a possibility, more precisely timed measures of participants' responses to specific elements in the stimuli were needed. The observation of overlapping terms used to describe responses to both music and language in the verbal labelling task in chapter 5 led to a fifth question of whether similar patterns would be observed in participants' use of the scales developed from those verbal labels in response to both music and language. The appearance of similar patterns between *harmony* and *semantics* seemed promising for support of incongruity-based models of music and language aesthetics. However, were a similar significant effect of *congruence* on aesthetic ratings to be found in both *harmony* and *semantics*, and were the scales shown to be used in similar ways across levels of *hrss* this would lead to a sixth question of whether the same mechanisms underlie this effect. Finally, the small effects of musical training warranted the question of whether differences would be seen between the responses of musical novices and highly trained musicians.

6.2.2 Choice of methods

Meyer's (2000) article claimed that incongruity-based models of aesthetics would only advance once methods which could determine participants' responses to specific elements in each stimulus were developed. These questions call for an experiment in which participants' responses to all four levels of *hrss* can be compared, in which behavioural measures could be captured and compared across *hrss* and in which participants' responses to specific elements in the stimuli could be monitored.

6.2.2.1 ERP methods

The ERP method consists in continuously recording variations in electric patterns at the scalp while participants perform a specific task, in this case, listening to stimuli. "Triggers" are sent from the software running the experiment to the software recording this continuous EEG data at specific moments. In the case of this experiment, these moments are the onset of the incongruity (the attack of the note in music, the onset of the first formant of the word in language). These triggers allow researchers to pinpoint the exact moments in the continuous recording at which point the element of interest occurred. Very little information can be derived from a single trial. The data occurring around the trigger are therefore averaged across trials in one condition (in this case, from 100ms before the trigger to 1300ms after the trigger), and compared across conditions.

This allows researchers to investigate the differences between average EEG patterns surrounding the element of interest in different conditions. In this experiment, the elements of interest are the incongruous chord or word and its resolution; the conditions are *normal*, *incongruous-resolved* and *incongruous-unresolved*. The amplitudes of the ERP waveforms are not informative in and of themselves. What is of interest is the difference between the patterns across different levels of *congruence*. In this experiment, the *normal* stimuli act as a baseline to which the other two conditions are compared in each level of *hrss*.

Part of this method's appeal is its high temporal resolution. Although methods such as Positron Emission Tomography and functional Magnetic Resonance Imaging have good temporal resolution, the delay involved in these methods of data collection means that responses cannot be associated with a specific moment in time during stimulus processing. Through the use of triggers, ERP methods do allow that precise pinpointing.

Furthermore, adopting this method allowed a comparability with a growing body of research carried out in this field, in which a majority of recent studies have used ERP methods. This enabled the present research to address unresolved issues arising from the systematic review, such as findings of both late positivities and late negativities elicited by harmonic incongruities (Patel et al., 1998; Koelsch, Gunter, et al., 2003), while building upon recent developments in the area of incongruity processing, notably in the area of syntactic aesthetically-purposed incongruities (Thierry et al., 2008).

6.2.2.2 The importance of behavioural data

One of the outcomes of the systematic review was a difficulty in the interpretation of ERP effects. While techniques for measuring participants' physiological responses have evolved to capture ever increasing amounts of data, theoretical advances in the area of incongruity-based aesthetics have been comparatively small (Juslin & Västfjäll, 2008). Partly to blame is the fact that many studies have left the relationship between the observed patterns and aesthetic-affective effects at a speculative level (Steinbeis et al., 2006). More systematic investigation is called for in both music and language, and theoretical advancements can only be reached when data are captured within a unified approach to minimise the need for speculations.

By combining behavioural data and ERP data, this experiment aimed to further the understanding of the effect of incongruities on different facets of participants' self-reported responses and establish the different contributors to aesthetic-affective responses. A broader research aim was to investigate the relationship between ERP effects and different aspects of participants' responses, with changes in ERP amplitude having previously been said to reflect oddity detection (Jentschke et al., 2005), processing difficulty (Kaan, Harris, Gibson, & Holcomb, 2000), syntactic integration (Patel, 2003), and, tentatively, to tension (Steinbeis et al., 2006) and aesthetic effects (Thierry et al., 2008).

6.2.3 Practical considerations

6.2.3.1 Recruitment

Although the effects of musical training reported in Experiment 1 were not significant, this lack of significance could have been due to the limited number of participants at the high end of the musical training spectrum. For this study, it was therefore decided to only recruit participants from either end of the musical training spectrum: novices with up to one year of extra-curricular musical training during childhood, and highly trained musicians who had achieved Grade 8, A-level music or an equivalent qualification.

Although these criteria do not provide a direct measure of participants' musicality, as do tests such as the Musical Ear Test (Wallentin, Nielson, Friis-Olivaruius, Vuust, & Vuust, 2010), they do provide a measure of whether participants have explicitly learnt the rules governing musical good form and western classical harmony. Furthermore, this is the measure adopted by researchers comparing musicians and non musicians in previous studies in this area (Bonnel et al., 2001; Koelsch et al., 2001; Jentschke et al., 2005; Miranda & Ullman, 2007; Steinbeis et al., 2006). Adopting this distinction between musicians and non musicians therefore allowed comparability with previous findings.

The studies reviewed while designing Experiment 3 tested a number of participants ranging from ten (Paller et al., 1992) to 64 (Miranda & Ullman, 2007), with the number of participants in each experimental group (musicians, non-musicianns) ranging from 10 to 32. Although the studies which inspired this research only tested 15 (Patel et al., 1998), 14 (Koelsch, Gunter, et al., 2003) and 12 participants in each group (Steinbeis et al., 2006), the average number of participants in each group across the studies in ERP research in music and language was 16.81. the wide variability in ERP data between participants, it is usually recommended that ERP experiments recruit approximately twenty participants (Luck, 2006).

A power analysis conducted in chapter 4 showed that 24 participants would be sufficient for a meaningful analysis of the behavioural data. Because ERP Piloting was carried out on non musicians, this resulted in a higher number of behavioural files (which were unchanged as a result of the piloting) than ERP files (for which the set-up procedure was optimised following piloting). Testing ceased when enough good ERP files had been gathered in both musicianship groups. Because of the necessity of discarding participants who presented less than 16 good data trials in each experimental condition due to signal to noise ratios (Luck, 2006), this led to a total of 30 non musicians and 25 non musicians in the final sample for the behavioural data, and 20 non musicians and between 20 and 17 musicians for the ERP data, remaining above the average number of participants tested in this field of research, and above the number of participants tested in studies central to the central research question of this thesis (*harmony*, 20, *rhythm*, 17, *semantics*, 17 and *semantics*, 18).

6.2.3.2 One big experiment

Because of the desire to investigate the possibility of similar mechanisms underlying the aesthetic effect of incongruities across music and language, an extensive experiment was constructed, in which participants were be tested with all four levels of *hrss*. ERP experiments require a minimum of sixteen trials per condition cell for the data to be informative due to signal to noise ratio (Luck, 2006). Because of the data loss incurred by participants moving, electrodes saturating or other sources of temporary noise, it is necessary to start with as many trials in each condition cell as is feasible within the design of the experiment.

Once the final stimulus set had been selected through the procedure detailed in Chapter 5, twenty four stimuli remained in each *congruence* condition. Piloting of the experiment showed that starting with 24 trials in each condition cell enabled a good retention of participants, allowing a loss of up to a third of the trials. Furthermore, in view of the duration of the stimuli, compared to the more typical visually presented stimuli used in many ERP experiments, it would not have been possible to test all four levels of *hrss* in one experiment had more trials been included in each condition cell. Thus, although more trials per condition cell would have been desirable, the present design allowed for the collecting of meaningful data, the analysis of which yielded significant effects of *congruence* similar to those reported in previous published studies in all four levels of *hrss* (see Chapter 8).

A distinction was made in the systematic review (Chapter 2) between rule-based incongruities (e.g., Patel et al., 1998), which are perceived as incongruous because they do not fit with implicitly learnt rules through exposure to music and language, and memory-based incongruities (e.g. Paller et al., 1992), which are incongruous by contrast with a participant's expectations of a familiar tune or language idiom through repeated exposure to that specific stimulus. Consequently, it was important that participants only heard one stimulus from each triplet (a triplet consists of a set of three stimuli, one *normal*, one *incongruous-resolved* and one *incongruous-unresolved* built around the same original aesthetically-purposed incongruity). Since there were 24 triplets, this meant that each participant could only hear 8 different stimuli in each *congruence* condition. Considering the concerns over the signal to noise ratio, the solution here was to repeat each stimulus three times.

The data collected here did not permit an investigation of the effect of this repetition on ERP patterns. These three repetitions were separated into three experimental blocks (A, B and C). This process is common in music ERP research (e.g. Steinbeis et al., 2006), and data from previous studies in psycholinguistics show no significant effect of repetition when the stimuli are repeated with a long lag between repetitions in a group of a similar age to the sample of this experiment (Rugg, Mark, Gilchrist, & Roberts, 1997). Furthermore, considering that many studies used a smaller number of repeated stimuli (e.g. six stimuli repeated six times each in the study by Steinbeis et al., 2006), and presented participants with different versions of each stimulus to create a sufficient number of trials (Jentschke et al., 2005; Steinbeis et al., 2006), the present study shows a greater degree of control over these potential confounds than previous research in this field. Furthermore, this repetition allowed three different sets of questions to be asked following each stimulus. Once again, the data here did not allow the investigation of whether the question which participants knew they were to answer after listening to the stimulus had an effect on participants' ERP data, but previous investigations have shown consistent ERP patterns under a variety of task conditions, and music and language ERP effects have been shown to be task-independent (Koelsch, Schroger, & Gunter, 2002; Magne et al., 2005; Ystad et al., 2007).

6.2.3.3 Separation between scales

Because of the mirror images observed between the ratings on the oddity scale and the ratings on the positive-to-negative response scale in Chapter 4 (see Figure 4.8), the ratings pertaining to oddity and to aesthetic appraisal needed to be separated in this experiment. This was done to investigate whether the correlation between the scales was a geniune effect of the experimental manipulation or a result of participants' responses to the oddity scale influencing their responses on the positive-to-negative scale. This was achieved by making use of these necessary repetitions and asking different questions each time each stimulus was repeated.

In block *A*, participants answered the following questions after hearing each stimulus:

1. Please use the mouse to indicate how odd you thought the stimulus sounded on the scale below — Scale from Completely normal to Very odd

2. In your opinion, the stimulus was... — Scale from Very bland to Very interesting

3. How confused or perplexed do you feel, having heard the stimulus? — Scale from Not at all to Very

In block *B*, participants answered the following questions after hearing each stimulus:

1. How aesthetically pleasing did you find the piece? — Scale from Not at all to Very

2. After listening to the whole piece, how do you feel? — Scale from Very relaxed to Very tense

In block *C*, participants answered the following questions after hearing each stimulus:

1. How stimulating did you find the stimulus? — Scale from Not at all to Very

2. How tense did you feel while you were listening to the piece? — Scale from Not at all to Very

6.3 Method

6.3.1 Participants

Participants were recruited from the University of Leeds community, through volunteer mailing lists, posters and, for musicians, special announcements in music lectures. non musicians were classified as participants with up to one year of extra-curricular musical training during childhood. Musicians were defined as participants having achieved either Grade 8 in a musical instrument or voice, and/or an A-level in music. All participants were naive to the hypotheses and to the experimental manipulation of the stimuli.

The selection criteria applied to all participants were as follows. Participants were all right handed, native speakers of British English, presenting no known language or hearing impairments and no skin conditions or wounds on their scalp. Participants having received brain surgery, having suffered from a neurological condition, or having had a severe brain injury requiring hospital treatment were excluded from the study, as were those taking any neurological or psychotropic medication such as anticonvulsants, antiepileptics, antidepressants, tranquilizers, neuroleptics and antipsychotics.

Before settling for the final set for the analysis, participants' accuracy in the calibration task was examined. The calibration task served to both verify participants' ability to click at the desired place on the screen, and determine the number of pixels between the right edge of the screen and the zero point of the rating scale. As in Experiment 1, the raw data scores were created by subtracting the number of pixels between the zero point on the scale and the right edge of the screen from the number of pixels between the point at which the participant clicked and the right edge of the screen. Thus, the data scores were dependent on a correct calibration. If a participant failed the calibration on up to two parts of the experiment, this was rectified systematically adding or subtracting the number of pixels by which the calibration was out to the raw scores provided in that part of the experiment. One participant (1117) failed the calibration by a large margin on several experimental blocks and showed a wide degree of variation between calibration points within each block (over 20 pixels), suggesting a poor accuracy in clicking at the appropriate place on the scale. This participant was removed from the data set out of concern that this lack of accuracy would translate into poor score reliability. A second participant (1108) was removed because she was not a native speaker of British English. The final sample comprised 30 non musicians (mean age 23.50 (sd = 10.01), 23 females, mean number of years of musical training 0.16 (sd = 0.37) and 25 musicians (mean age 23.56 (sd = 9.50), 13 females, mean number of years of musical training 11.91 (sd = 3.31)).

Because of the necessity of discarding participants with less than 16 good trials in each condition cell for the ERP analysis, the sample for the ERP analysis presented in Chapter 8 consisted in 20 non musicians (15 females, mean age = 23.00, sd = 10.43, mean years musical training = 0.15, sd=0.36) and 20 musicians (14 females, mean age = 23.40, sd = 9.76, mean years of musical training = 9.77, sd = 3.19). Since the number of partipants necessary for the behavioural data was 24 in each group, but the number necessary for the ERP data was only 20 in each group, it was decided to keep all the participants in the behavioural set, regardless of their ERP data. However, when the ERP data and the behavioural data were statistically compared (see last section of Chapter 8), only participants with sufficient ERP data were retained.

6.3.2 EEG data recording

The EEG recording used the NeuroScan 4.3 Acquire software and a Synamps2 amplifier with a 64-channel Ag-AgCl QuikCell electrodes cap placed according to the Extended International 10-20 system and two additional electrodes placed on the mastoids (see Figure 6.1). Vertical and horizontal electrooculograms were recorded by placing one electrode on the outer canthi of both eyes, one above the right eye and one below the right eye to monitor eye movements. The ground electrode was positioned between FPz and Fz. Data were recorded using a central reference positioned between Cz and CPz. The continuous EEG data were sampled at 1000Hz and filtered online using a 200Hz low-pass filter.

Between each part of the experiment, the experimenter checked the conductance of the electrodes and remoistened any which had lost conductance. Since the QuikCell procedure allows good quality data with both high and low impedances, this action was

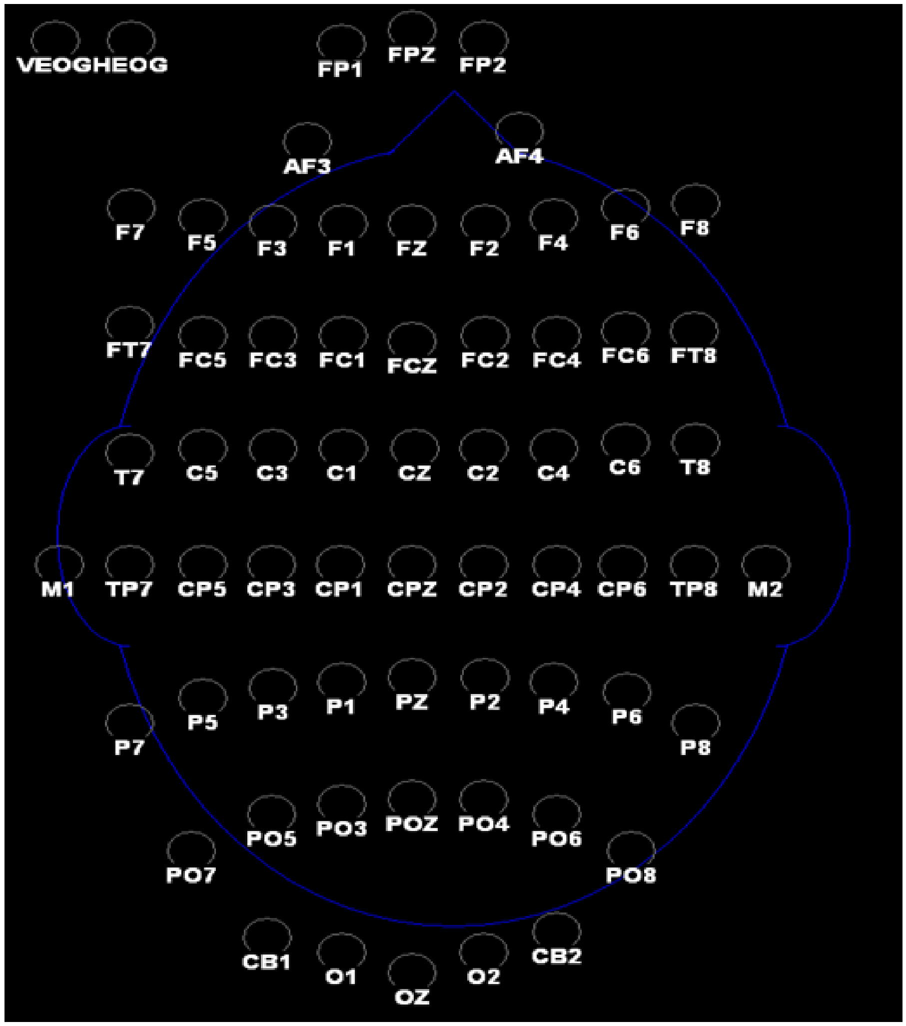

Figure 6.1: Electrode placements used in Experiment 3

normally taken only between each of the blocks, but this systematic checking combined with online monitoring of the EEG data allowed failing electrodes to be salvaged mid block if necessary.

6.3.3 EEG data pre-processing

EEG data were analysed offline using NeuroScan 4.3 Edit software. The procedures described in this section were carried out for each of the 12 files within each participant's data set, to ensure that no slight changes due to a remoistening of the electrodes affected the efficiency of this part of the data preparation.

Data were re-referenced offline to the average of the left and right mastoid electrodes.

They were then filtered using a band-pass filter of 0.1 (high-pass filter) to 30Hz (low-pass filter) with a slope of 24dB/octave. The continuous data were visually inspected and segments were rejected if they appeared to be very noisy or saturated. Eyeblink artifacts were corrected using NeuroScan ocular artefact rejection based on a minimum of 32 blinks per participant. EEG data were epoched and ERPs formed timelocked to the onset of the incongruity and the resolution in the music stimuli, and timelocked to the incongruity and each word after the incongruity in the language stimuli, from 100ms before the onset of the element of interest to 1300ms after the onset of the element of interest. Epochs were excluded if the amplitude exceeded +/- 75μV on any channel. Data were smoothed over five points and baseline corrected using the pre-stimulus interval (-100ms to stimulus onset). Non-rejected epochs were averaged from 0 to 1300ms after the onset of the element of interest with a baseline from -100ms to 0ms.

6.3.4 Stimuli, design and procedure

The participants were tested using a computer running the software *E-Prime* version 1.2 in a small room in the department's EEG laboratory visible to the experimenter via CCTV. Participants were fitted with the EEG cap and appropriate levels of conductance for the electrodes were established before the experiment began. Closed noise cancelling digital stereo headphones, placed over the EEG cap, were used to play the stimuli. During listening, a fixation point was provided in the form of an asterisc on the screen.

The stimuli were the final set selected through the norming procedure. The total set consisted of twenty four stimuli in each *congruence* condition (*normal*, *incongruous-resolved* and *incongruous-unresolved*) and in each level of *hrss* (*harmony*, *rhythm*, *semantics* and *syntax*) — a total of 288 (3 x 4 x 24) stimuli. Each participant heard only one stimulus from each triplet within each level of *hrss*, meaning that each participant heard a total of 96 (8 (stimuli) x 3 (*congruence* condition) x 4 (*hrss*)) different stimuli. As discussed in the Rationale section of this chapter, these were repeated three times to build up the number of trials necessary for ERP analyses and to allow three different sets of questions to be asked about each stimulus.

This repetition and the separation of levels of *hrss* within each question block meant that the experiment comprised twelve parts (3 question blocks x 4 levels of *hrss*). The order of the three blocks was randomised across participants, as was the order of the four parts within each block, while maintaining a systematic alteration of language and music. For example, participant 1 started with Block A, answering questions concerning oddity, interest and confusion, in the order $harmony-semantics-rhythm-syntax$. Once this block was completed, Block B began, with questions concerning aesthetic pleasantness and post-stimulus tension, with the same order of parts as in Block A. This participant's final block was C, with questions concerning stimulation and tension during listening, with the parts for each level of *hrss* in the same order again. Participant 2 started with Block B, hearing all 24 stimuli from each level of *hrss* separately in the order $semantics-rhythm-syntax-harmony$ then performed Block C and finally Block A with the same order of *hrss* within each block. Participant 3 started with Block C in the order $rhythm - semantics - harmony - syntax$, before proceeding to Block B and Block A, and so forth. The experimental software randomised the stimuli presented within each part of the experiment, controlling for order effects both within and across participants.

The scales were those described in the previous section, and were answered in the same manner as the scales used in the first rating experiment (Chapter 4). The entire experiment lasted approximately three hours, including EEG set-up and short breaks between each of the blocks of the experiment. All participants received £10 or 14 course credits for their participation.

6.4 Conclusion

The goal of this complex experiment was to arrive at a model showing the different contributors to heightened aesthetic-affective responses to music and language, which would take into account both the level of incongruity of the stimulus itself and the effect of that incongruous nature on different facets of participants' responses. The methods detailed in this chapter led to a vast amount of data. The analysis was therefore split over two chapters, with the first reporting the findings from the behavioural data (Chapter 7) and the second reporting the ERP effects and their relationship to the behavioural data (Chapter 8).

Chapter 7

Mildly odd and very pleasant: Behavioural data surrounding the processing of aesthetically-purposed incongruities

7.1 Introduction

This chapter presents the analysis and discussion of behavioural data gathered in Experiment 3. The first section examines the effect of *congruence* (*normal*, *incongruous-resolved* and *incongruous-unresolved*) on each of the scales used in this experiment, both across all four levels of *hrss* and within each level of *hrss* (*harmony*, *rhythm*, *semantics* and *syntax*). Within this section, the reliability of the *congruence* manipulation detailed in Chapter 3 and the the validity of the norming procedure presented in Chapter 5 are tested. The second section aims to investigate the relationships between the data from these scales, to build a model of how the different aspects of participants' responses to incongruities contribute to participants' aesthetic appreciation of stimuli across domains. Results are discussed alongside the analysis.

7.2 Preparing the data set

A first category of outliers consisted in data points for which participants had failed to click within the area delimited by the scale on the screen. A grace area of 10 pixels was allowed. Data points which fell outside the scale by more than 10 pixels were replaced by a missing value. Following this "erroneous click idenfication" phase, data were prepared and analysed following the procedures advocated by Tabachnick and Fidell (2007). Outliers were identified based upon their z-scores and degrees of "'disconnectedness" within each condition cell (separating the data by *hrss* and by *congruence*). Individual data points were replaced by missing values if and only if their z-score was equal to or above 2.58 *and* their representation on the box-plot for that condition cell showed disconnectedness from the whiskers of the boxplot. This procedure resulted in the rejection of only 1% of all data points.

The file was then split by participant and by *hrss* (*harmony, rhythm, semantics, syntax*) to compute z-scores, in order to maximise the use of the rating scales by each participant, using the same approach as in the initial rating study (see chapter 4 for a more detailed justification of this approach). Since the nature of ERP data only allows a by-subjects analysis of the effect of *congruence*, it was decided to then average the scores thus obtained within each cell for each participant, to allow a by-subjects analysis of these behavioural data. The data presented in this chapter are therefore the means of and standard errors of the means of the data points within each level of *congruence* within each level of *hrss* within each subject. This approach resulted in the same outcome as if missing values had been replaced by the mean value of each experimental cell for each participant.

7.3 Effects of *congruence* on each of the rating scales

For these analyses, 2 (musicians, non musicians) x 3 (normal, incongruous-resolved, incongruous-unresolved) x 4 (harmony, rhythm, semantics, syntax) ANOVA were carried out on the mean z-transformed rating scores by *congruence* and per *hrss* in a by-subjects analysis. As in Chapter 4, significant interactions were followed up with simple effects analyses (also referred to as *simple main effects*). These were done, in line with the guidelines of Tabachnick and Fidell (2007), by selecting only the relevant cases (i.e. musicians or non musicians to examine the effect of the between-groups variable) or relevant experiment cells (i.e. only data relevant to *harmony* stimuli for a repeated-measures effect). Significant simple main effects were followed up by simple contrasts where theory allowed planned contrasts and by a Bonferroni pairwise comparison where a more exploratory approach was needed. The order in which the analyses are presented does not reflect the order in which participants used the rating scales, as this order varied across participants.

7.3.1 Oddity

The first scale for which the data were analysed was "Please use the scale below to indicate how odd you thought the stimulus sounded" rated from "very odd" to "completely normal". This question was the same as the one asked in the first rating study, and its aim here was to provide a validation for the stimulus selection procedure described in Chapter 5. The data pooled across both participant groups are presented in Figure 7.1.

Since the experimental manipulation of the stimuli was designed to create stimuli which would reliably be rated as normal, slightly odd and very odd, following the norming procedure described in Chapter 5, a clear one-tailed hypothesis could be used to direct planned contrasts: it was predicted that the following pattern would occur in all four levels of *hrss*: *normal* < *incongruous-resolved* < *incongruous-unresolved*. This one-tailed pattern was assessed using a *Repeated* contrast analysis, comparing the *normal* condition to the *incongruous-resolved* condition, and the *incongruous-resolved* condition to the *incongruous-unresolved* condition.

The overall ANOVA revealed a significant main effect for *congruence* ($F(2, 100) =$

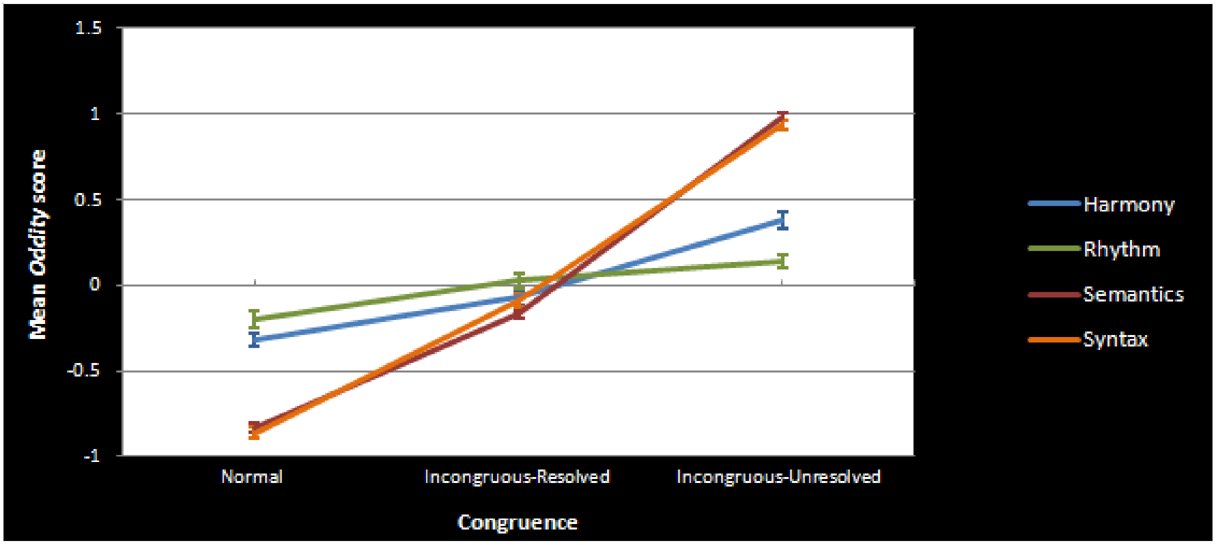

Figure 7.1: *Oddity* ratings by *hrss* and by *congruence*

$586.56, p < 0.001, \eta^2 = .92$). Planned contrasts showed a significant increase in oddity ratings from *normal* to *incongruous-resolved* ($F(1, 50) = 158.85, p < 0.001, \eta^2 = 0.76$) as well as a significant increase in oddity ratings from *incongruous-resolved* to *incongruous-unresolved* ($F(1, 50) = 458.10, p < 0.001, \eta^2 = 0.90$). A significant *hrss* x *congruence* interaction effect was also found ($F(6, 300) = 77.21, p < 0.001, \eta^2 = .61$), reflecting the lower incline in the linear trend of the music stimuli compared to that observed in the language stimuli. A significant three-way interaction (*musicianship* x *hrss* x *congruence*: $F(6, 300) = 2.78, p = 0.01, \eta^2 = 0.05$) called for simple main effects analyses within each level of *hrss* and *musicianship*.

Within *harmony*, a significant main effect for condition was found ($F(2, 106) = 45.36, p < 0.001, \eta^2 = 0.46$). The same simple contrasts were applied, revealing a strongly significant increase in oddity rating at each level of the *congruence* variable: *normal* to *incongruous-resolved* ($F(1, 53) = 13.02, p < 0.001, \eta^2 = 0.20$) and *incongruous-resolved* to *incongrous-unresolved* ($F(1, 53) = 34.53, p < 0.001, \eta^2 = 0.38$). As could be expected from Figure 7.2, a significant *congruence* x *musicianship* interaction effect ($F(2, 104) = 5.12, p = 0.01, \eta^2 = 0.09$) was also found. This effect was followed up by selecting only the cases relevant to each level of the interacting variable (musicianship).

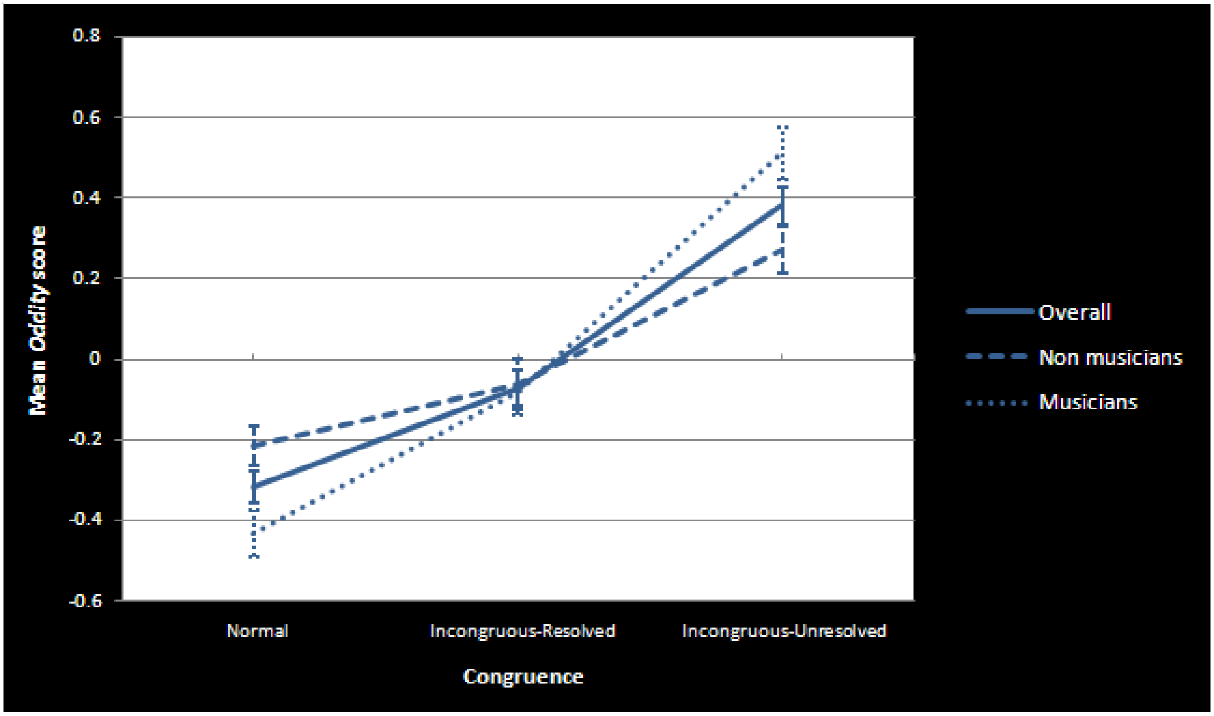

Figure 7.2: *Oddity* ratings by *congruence* within *harmony* split by *musicianship*

These analyses revealed a significant main effect for *congruence* in both groups ($F(2,56) = 12.82, p < 0.001, \eta^2 = 0.31$ in non musicians, $F(2,48) = 41.20, p < 0.001, \eta^2 = 0.63$ in musicians), with a stronger effect in musicians than in non musicians. Simple contrasts revealed that while the musicians showed a significant increase at each level of the *congruence* variable (all $p \leq 0.001$), the increase in rating from *normal* to *incongruous-resolved* in non musicians just failed to reach significance ($F(1) = 2.63, p = 0.06, \eta^2 = 0.09$). *Incongrous-unresolved* stimuli were however rated as significantly more odd than *incongruous-resolved* stimuli by non musicians ($p = 0.006$).

Within *rhythm*, the same approach to analysis revealed a significant main effect for *congruence* ($F(2,106) = 10.91, p < 0.001, \eta^2 = 0.17$). A significant increase was noted between each level of *congruence* in the one-tailed simple contrasts analysis: *normal* to *incongruous-resolved* ($F(1,53) = 10.82, p = 0.002, \eta^2 = 0.17$) and *incongruous-resolved* to *incongruous-unresolved* ($F(1,53) = 2.75, p = 0.05, \eta^2 = 0.05$). Once more, as suggested by Figure 7.3 a significant *congruence* x *musicianship* interaction was found ($F(2,106) = 4.33, p = 0.02, \eta^2 = 08$).

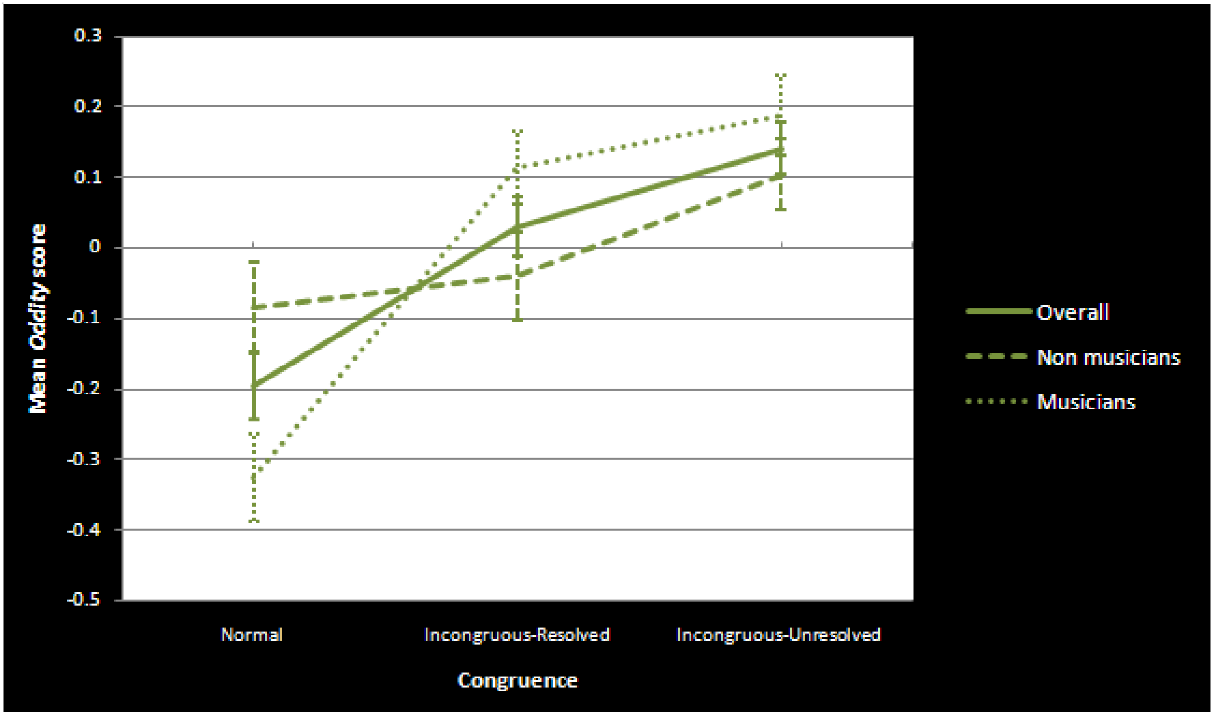

Figure 7.3: *Oddity* ratings by *congruence* within *rhythm* split by *musicianship*

A simple effect analysis within each level of musicianship revealed that a main effect for *congruence* was only seen in musicians ($F(2,48) = 16.12, p < 0.001, \eta^2 = 0.40$). Looking at the trends in these groups separately revealed a striking difference between how both groups responded to the *incongruous-resolved* condition which could explain the smaller effect size noted in the second contrast. As shown in Figure 7.3, though trends were in the expected direction in both groups, non musicians rated the *incongruous-resolved* condition as closest to the *normal* condition and much less odd than the *incongruous-unresolved* condition, while musicians rated it as closer to the *incongruous-unresolved* condition and significantly more odd than the *normal* condition ($F(1,24) = 20.67, p < 0.001, \eta^2 = 0.46$).

Within *semantics*, the significant main effect for *congruence* remained ($F(2,106) = 913.05, p < 0.001, \eta^2 = 0.95$), with a significant increase in oddity ratings at each level of *congruence* (all contrasts $p < 0.001, \eta^2 > 0.80$). The *congruence* x *musicianship* interaction failed to reach significance ($F(2,104) = 2.61, p = 0.08, \eta^2 = 0.05$).

Finally, the analyses were applied to *syntax*, where a significant main effect for *congruence* was found ($F(2, 106) = 489.89, p < 0.001, \eta^2 = 0.90$). Once more, simple contrasts showed a significant increase at each step (all $p < 0.001, \eta^2 > 0.70$). No significant *congruence* x *musicianship* effect was found ($F(2, 104) = 0.38, p = 0.69, \eta^2 = 0.007$).

In summary, all four levels of *hrss* showed a significant increase in oddity ratings from *normal* to *incongruous-resolved* to *incongruous-unresolved*. Although the trends were of differing strength across *hrss*, all differences were significant, confirming the validity of the experimental manipulation and the validity of the stimulus norming procedure described in chapter 5. This section allows us to advance with confidence that the stimulus set is meeting its objectives when all participants are pooled together. The differences between musicians and non musicians call for caution in the generalisability of the effect of *congruence* across groups, especially in the rhythm stimuli. These difference could be explained by musicians being more used to maintaining the meter while listening to ambiguous rhythmic sections. Nonetheless, the stimuli can be confirmed as validated for the purpose of this research project.

7.3.2 Aesthetic pleasantness

These data were collected using the question "How aesthetically pleasing was the stimulus as a whole?" rated from "not at all" to "very". The theories which provided the impetus for this thesis hold that stimuli containing an incongruity which resolves in a way which allows the integration of the stimulus into a meaningful whole are more pleasant than stimuli which do not diverge from the expected norm. Furthermore, in view of the striving for integration which underlies the beneficial effects of a good resolution, these theories would also suggest that a stimulus which does not resolve should be less pleasant than a stimulus containing no incongruities. The planned contrasts for this analysis are therefore *simple* contrasts, comparing each of the incongruous levels of *congruence* separately to the *normal* level. The one-tailed predictions are as follows. *Incongruous-resolved stimuli* will receive a significantly higher aesthetic rating than *normal* stimuli. *Incongruous-unresolved* stimuli will receive a significantly lower aesthetic rating than *normal* stimuli.

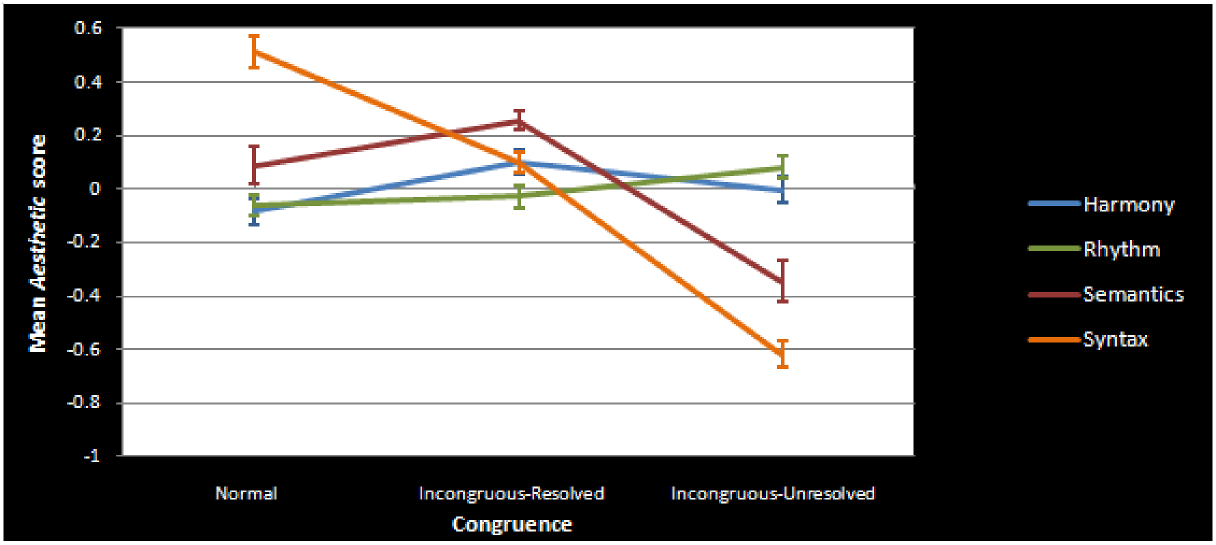

Figure 7.4: *Aesthetics* ratings by *hrss* and by *congruence*

The 2 x 3 x 4 ANOVA carried out on the aesthetic rating data shown in Figure 7.4 demonstrated a significant main effect for *congruence* ($F(2, 100) = 26.35, p < 0.001, \eta^2 = 0.35$), with no significant increase in mean ratings between *normal* and *incongruous-resolved* ($F(1, 50) = 0.24, p = 0.61, \eta^2 = 0.005$) but a significant decrease in mean ratings from *normal* to *incongruous-unresolved* ($F(1, 50) = 55.25, p < 0.001, \eta^2 = 0.53$). A significant *hrss* x *congruence* interaction was also found ($F(6, 300) = 29.21, p < 0.001, \eta^2 = 0.37$) but no significant three-way interaction. Nonetheless, since the trends observed in *harmony* and *semantics* show differences between the two groups (see Figures 7.5 and 7.6), the two levels of musicianship were analysed separately in *harmony* and *semantics*.

Within *harmony*, the main effect for condition failed to reach significance when both groups were analysed together ($F(2, 106) = 2.23, p = 0.11, \eta^2 = 0.04$) despite a significant increase in rating between *normal* and *incongruous-resolved* stimuli in support of the one-tailed hypothesis ($F(1, 54) = 5.24, p = 0.02, \eta^2 = 0.09$). Although the *congruence* x *musicianship* effect was not significant ($F(2, 106) = 0.57, p = 0.57, \eta^2 = 0.01$), the differences in trends between musicians and non musicians shown in Figure 7.5

called for an investigation of *congruence* within each group separately.

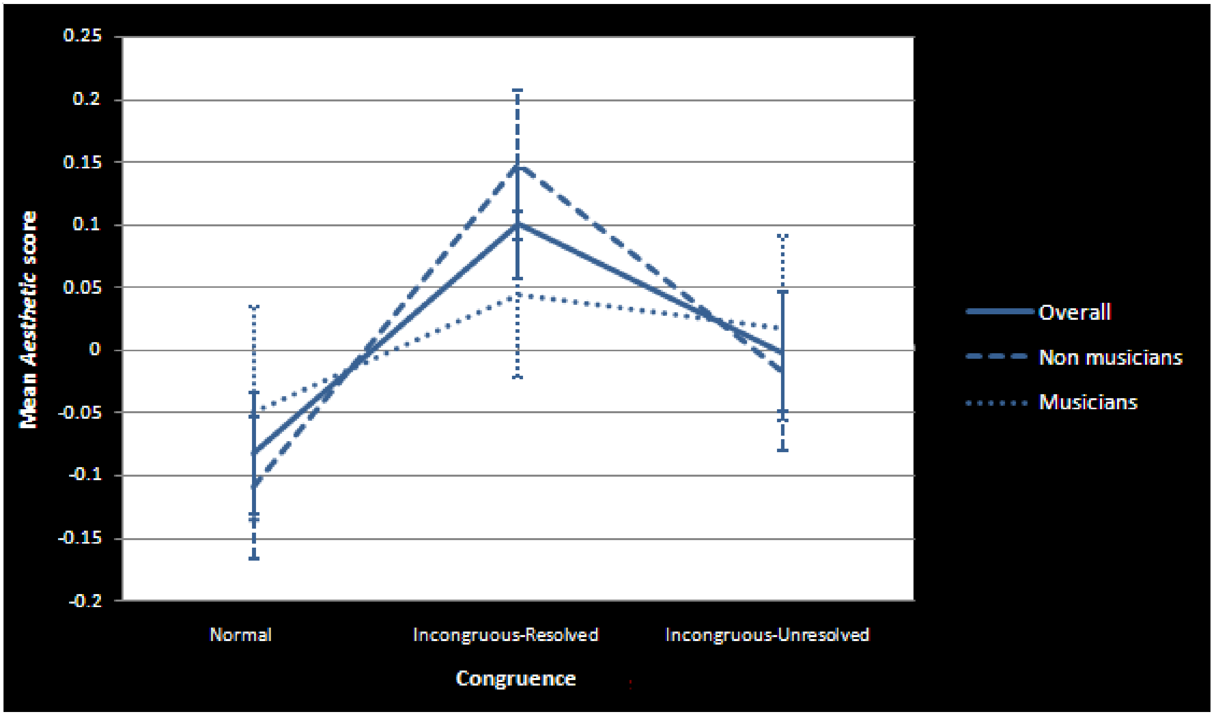

Figure 7.5: *Aesthetics* ratings by *congruence* within *harmony* split by *musicianship*

These further simple effects analyses revealed that within *harmony*, the main effect for *congruence* was significant only in non musicians ($F(2,58) = 3.16, p = 0.05, \eta^2 = 0.10$, musicians $F(2,48) = 0.28, p = 0.76, \eta^2 = 0.01$). A simple contrast analysis following the significant effect in non musicians showed that *incongruous-resolved* stimuli received a significantly higher rating than *normal* stimuli ($F(1,29) = 7.23, p = 0.006, \eta^2 = 0.20$), in support of theories of incongruity-based aesthetics in *harmony*.

Within *rhythm* the main effect for *congruence* was not significant ($F(2,108) = 2.12, p = 0.125, \eta^2 = 0.04$) and since the trend was in the unexpected direction (*incongruous-unresolved* stimuli were rated more aesthetically pleasant than *normal* stimuli ($F(1,54) = 2.27, p = 0.03, \eta^2 = 0.07$)), the significance of the planned contrast analyses could not be applied meaningfully to this set of data. Bonferroni post-hoc analyses revealed no significant differences in pairwise comparisons between all three levels of *congruence*. No significant *congruence* x *musicianship* effect was found ($\mathrm{F}(2,2,106) = 0.05, p = 0.95, \eta^2 = 0.001$).

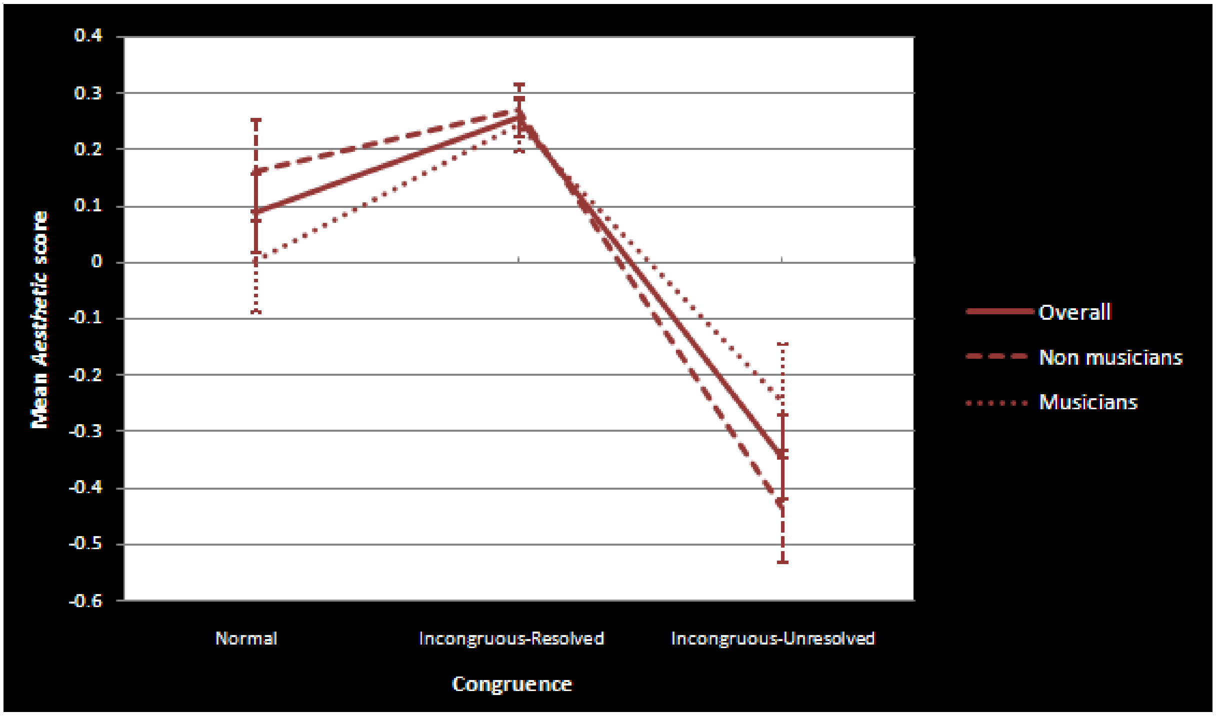

Figure 7.6: *Aesthetics* ratings by *congruence* within *semantics* split by *musicianship*

Within *semantics*, a strongly significant main effect for *congruence* was found ($F(2, 102) = 15.79, p < 0.001, \eta^2 = 0.24$). Simple contrasts showed that *incongruous-resolved* stimuli were rated significantly more aesthetically pleasant than *normal* stimuli ($F(1, 52) = 4.34, p = 0.02, \eta^2 = 0.08$), which in turn received a significantly higher rating than *incongruous-unresolved* stimuli ($F(1, 52) = 9.30, p = 0.004, \eta^2 = 0.15$). Though no significant *congruence* x *musicianship* interaction was found ($F(2, 102) = 1.29, p = 0.27.\eta^2 = 0.03$), the differences in trends between musicians and non musicians (see Figure 7.6) prompted further simple effects analyses as in the case of the *harmony* data.

These simple effect analyses revealed revealed a significant main effect for *congruence* within both groups (musicians, $F(2, 48) = 4.25, p = 0.02, \eta^2 = 0.15$, non musicians $F(2, 54) = 14.28, p < 0.001, \eta^2 = 0.35$). The planned contrasts analyses showed a significant increase between *normal* and *incongruous-resolved*, predicted by incongruity-based theories of aesthetics, in musicians only ($F(1, 24) = 3.34, p = 0.04, \eta^2 = 0.12$, non musicians, $F(2, 24) = 1.12, p = 0.30, \eta^2 = 0.04$). A significant decrease between *normal* and *incongruous-unresolved* was found in non musicians only ($F(1, 27) = 10.48, p = 0.001, \eta^2 = 0.28$, musicians, $F(1, 24) = 1.32, p = 0.24, \eta^2 = 0.05$).

Finally, within *syntax* a significant main effect for *congruence* was revealed ($F(2, 106) = 82.18, p < 0.001, \eta^2 = 0.61$). In both musicians and non musicians, the trend was a decrease from *normal* to *incongruous-resolved* and a further decrease from *incongruous-resolved* to *incongruous-unresolved*, calling for post-hoc analyses rather than the planned contrasts. A Bonferroni analysis revealed significant differences in all pairwise comparisons between the three levels of *congruence*, meaning that the decrease in aesthetic ratings from *normal* to *incongruous-resolved* and from *incongruous-resolved* to *incongruous-unresolved* were both significant (all differences, $p < 0.001$). No significant *congruence* x *musicianship* interaction effect was found ($F(2, 104) = 0.87, p = 0.40, \eta^2 = 0.02$).

In summary, support for incongruity-based theories of aesthetics was found only in *harmony* and *semantics*. Though small, these effects provide the first set of direct evidence in support of theories of incongruity-based aesthetics in harmony and semantics using a theoretically-based and empirically validated set of auditory stimuli allowing for the investigation of the effect of both novelty and resolution. Splitting the analysis by musicianship group, non musicians were the only ones to show a significant increase in aesthetic ratings for *incongruous-resolved* harmonic stimuli, while musicians only showed the significant increase in aesthetic ratings for *incongruous-resolved* semantic stimuli. These differences are further explored in chapter 8 in the light of significant differences between groups in ERP patterns.

The main aim in collecting responses on the next set of scales was to see the different contributors to aesthetic pleasantness through a regression analysis. However, for the sake of completeness, these responses are analysed in the same way as the previous two scales. The exploratory nature of this analysis called for post hoc pairwise comparisons. Despite the risk of power loss, a Bonferroni analysis was chosen for consistency with the exploratory analysis carried out in Chapter 4.

7.3.3 Interest

These data were collected using the question "In your opinion, the stimulus was..." answered on a scale ranging from "very bland" to "very interesting". The 2 x 3 x 4 ANOVA revealed a significant main effect for *congruence* ($F(2, 100) = 27.34, p < 0.001, \eta^2 = 0.35$), showing a significant difference between all levels of *congruence* in a Bonferroni post-hoc analysis: a significant increase in mean interest scores was seen from *normal* to *incongruous-resolved* and again from *incongruous-resolved* to *incongruous-unresolved*. A significant *hrss* x *congruence* interaction ($F(6, 300) = 7.57, p < 0.001, \eta^2 = 0.13$) was also found, calling for simple main effects analyses, but no significant three-way interaction with musicianship was found ($F(6, 600) = 0.86, p = 0.53, \eta^2 = 0.02$).

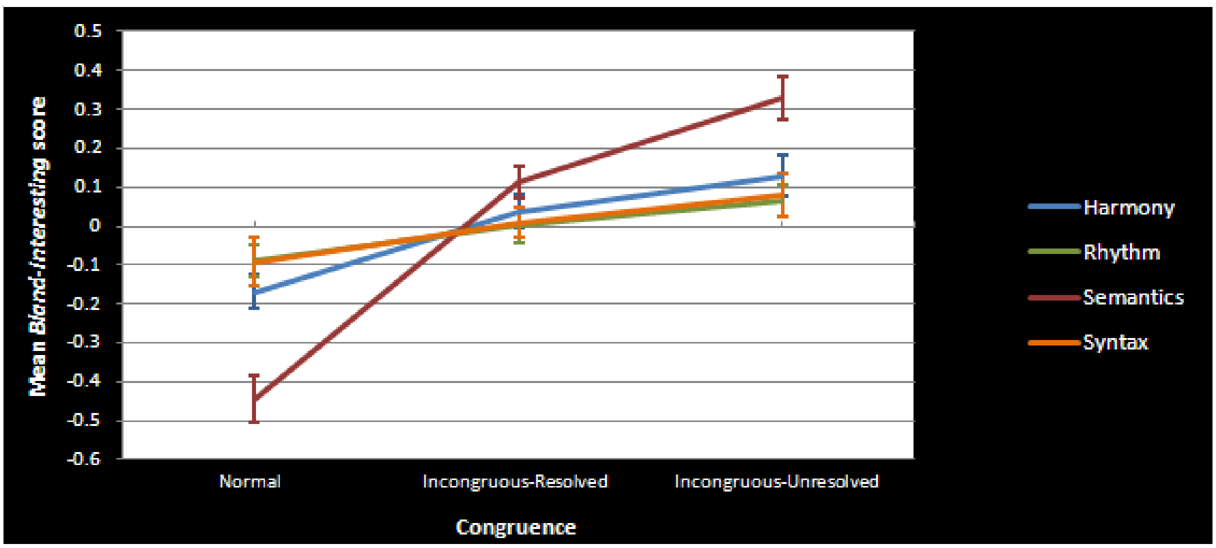

Figure 7.7: *Bland-Interesting* ratings by *hrss* and by *congruence*

The simple effects analysis within *harmony* showed a significant main effect for *congruence* ($F(2, 104) = 7.10, p = 0.001, \eta^2 = 0.12$), with the *normal* condition receiving a significantly lower mean rating of interest than the other two *congruence* conditions (both differences $p \leq 0.01$). Though the *congruence* x *musicianship* interaction failed to reach significance ($F(2, 104) = 2.27, p = 0.11, \eta^2 = 0.04$), the trends shown in Figure 7.8 prompted a separate analysis for each level of musicianship.

When splitting the data by musicianship, no significant main effect for *congruence* was seen in non musicians ($F(2, 56) = 1.32, p = 0.28, \eta^2 = 0.05$) in *harmony*, but a

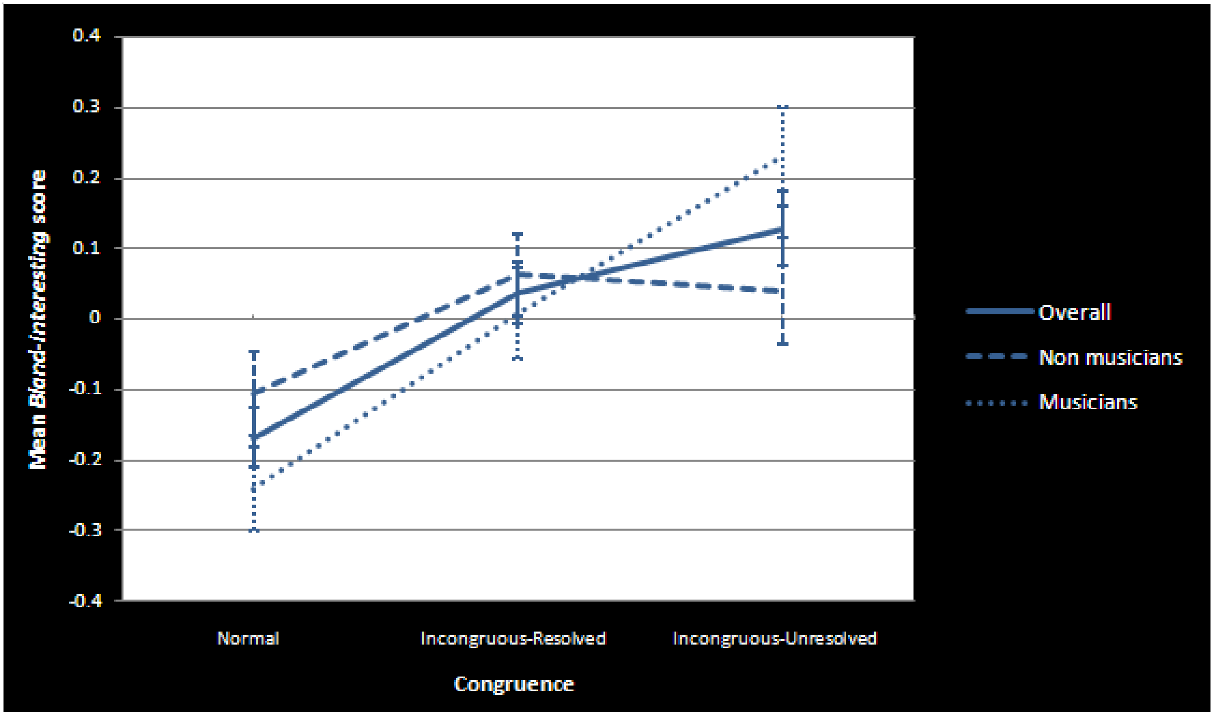

Figure 7.8: *Interest* ratings by *congruence* within *harmony* split by *musicianship*

strongly significant effect was found in musicians ($F(2,48) = 8.80, p = 0.001, \eta^2 = 0.268$). Pairwise comparisons in the musician data revealed that the increase in ratings from *normal* to *incongruous-resolved* failed narrowly to reach significance ($p = 0.07$) but that *incongruous-unresolved* stimuli were rated significantly more interesting than *normal* stimuli ($p = 0.001$). No significant difference was seen between the *incongruous-resolved* and *incongruous-unresolved* conditions ($p = 0.24$).

Within *rhythm*, neither a significant main effect for *congruence* ($F(2,106) = 2.11, p = 0.13, \eta^2 = 0.04$) nor a significant interaction with musicianship ($F(2,106) = 0.16, p = 0.85, \eta^2 = 0.003$) were found although the linear trend (*normal* $<$ *incongruous-resolved* $<$ *incongruous-unresolved*) was significant ($F(1,54) = 4.59, p = 0.04, \eta^2 = 0.08$).

Within *semantics*, a significant main effect for *congruence* was found ($F(1,104) = 36.73, p < 0.001, \eta^2 = 0.41$) with a significant increase at each step of the linear trend (*normal* $<$ *incongruous-resolved* $<$ *incongruous-unresolved*, differences respectively $p < 0.001$ and $p = 0.01$). Although no significant interaction with musicianship was observed ($F(2,104) = 1.08, p = 0.34, \eta^2 = 0.02$), as in the case of *harmony*, the

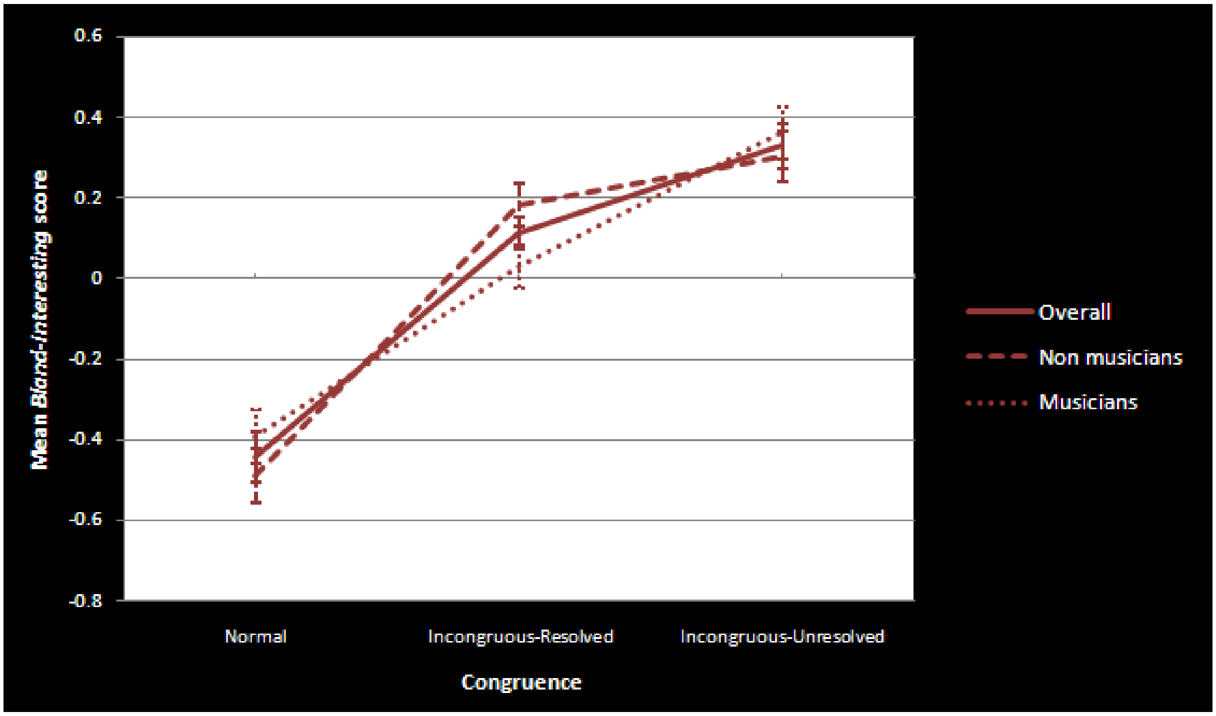

Figure 7.9: *Interest* ratings by *congruence* within *semantics* split by *musicianship*

data trends suggested different effect sizes in the two groups (see Figure 7.9).

Indeed, when splitting the data by musicianship, a significant main effect for *congruence* was seen in both non musicians and musicians (respectively $F(2, 56) = 32.15, p < 0.001, \eta^2 = 0.53$ and $F(2, 48) = 11.53, p < 0.001, \eta^2 = 0.33$), but although both groups rated *incongruous-resolved* stimuli as significantly more interesting than *normal* stimuli (respectively $p < 0.001$ and $p = 0.03$), only musicians rated the *incongruous-unresolved* stimuli significantly more interesting than the *incongruous-resolved* stimuli (non musicians $p = 0.64$, musicians $p = 0.03$).

Finally, within *syntax*, neither a main effect for *congruence* ($\mathrm{F}(2, 104) = 1.83, p = 0.17, \eta^2 = 0.03$) nor a significant *congruence* x *musicianship* effect ($F(2, 104) = 0.11, p = 0.86, \eta^2 = 0.002$) were found.

The *interest* scale therefore only showed a significant main effect for *congruence* within *harmony* and *semantics*. In both these levels of *hrss*, a linear trend was noted, but trends differed between musicianship groups. In *harmony*, a significant effect of *congruence* was seen only in musicians, *incongruous-unresolved* stimuli were rated as significantly more interesting than *normal* stimuli. In *semantics*, only the musicians rated the *incongruous-unresolved* stimulus as significantly more interesting than the *incongruous-resolved* stimulus. It therefore appears that the lack of resolution rendered an incongruous stimulus more interesting to the musicians but no more interesting to the non musicians.

7.3.4 Confusion

These data were collected on the following scale: "How confused or perplexed do you feel having heard the whole stimulus?", with anchors "not at all" to "very". The overall ANOVA for confusion showed a significant main effect for *congruence* ($F(2, 100) = 358.50, p < 0.001, \eta^2 = 0.88$), a significant interaction of *hrss* with *congruence* ($F(6, 300) = 65.01, p < 0.001, \eta^2 = 0.57$), of *congruence* with *musicianship* ($F(2, 100) = 7.05, p = 0.002, \eta^2 = 0.12$) but no significant three-way interaction. The significant effect for *congruence* was followed up with pairwise comparisons, which showed significant differences in all comparisons, summarised by a significant increase in confusion ratings (indicating a greater degree of confusion) from *normal* to *incongruous-resolved* and from *incongrous-resolved* to *incongruous-unresolved*.

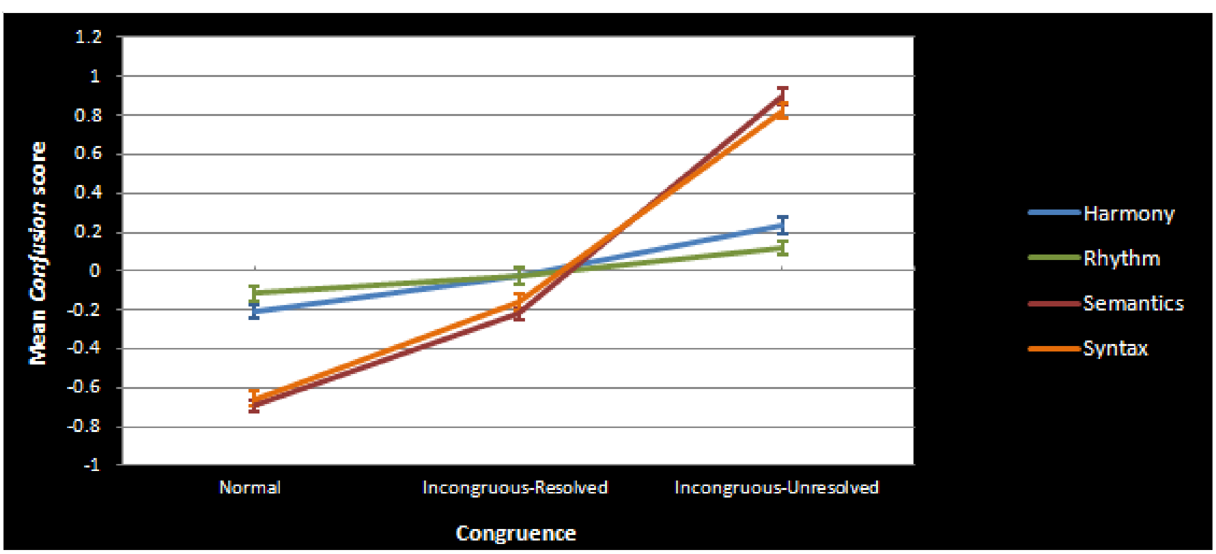

Figure 7.10: *Confusion* ratings by *hrss* and by *congruence*

Within *harmony*, the simple effects analysis revealed a significant main effect for *congruence* ($F(2,106) = 20.28, p < 0.001, \eta^2 = 0.28$) with a significant increase in confusion rating at each step of the linear trend (differences respectively $p = 0.008$ and $p = 0.003$). As could be expected from the trends shown in Figure 7.11, a significant *congruence* x *musicianship* interaction was also seen ($F(2,104) = 5.04, p = 0.01, \eta^2 = 0.09$).

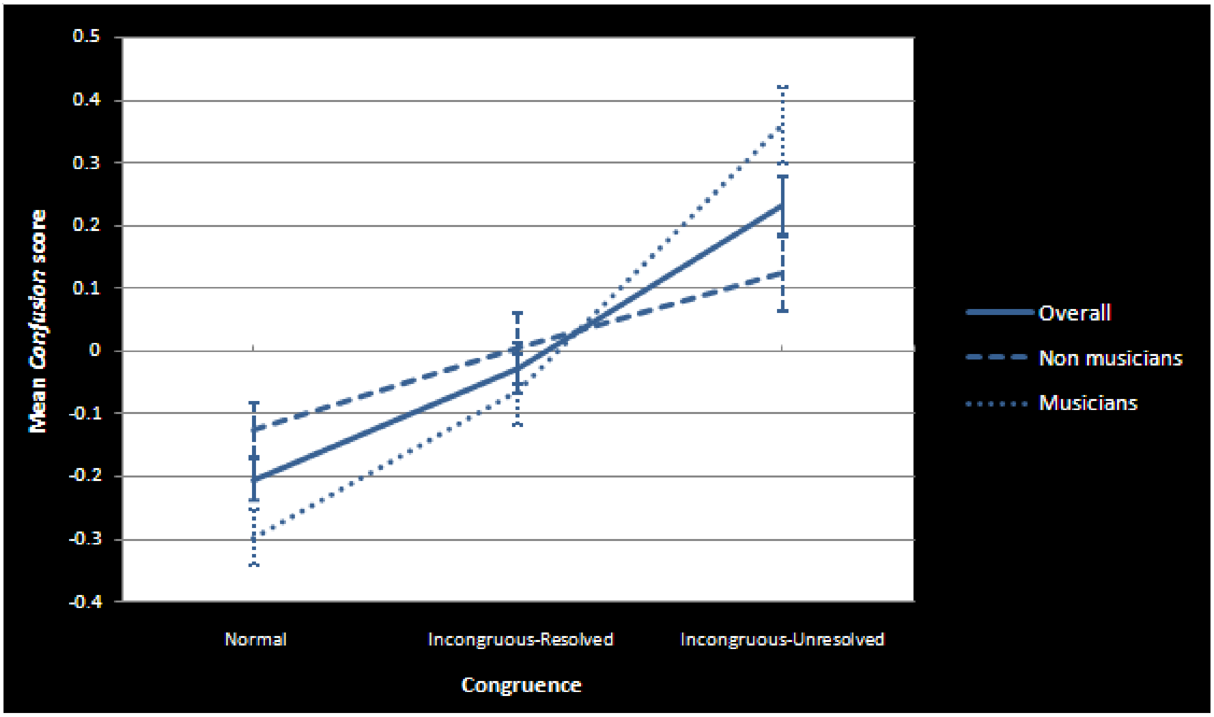

Figure 7.11: *Confusion* ratings by *congruence* within *harmony* split by *musicianship*

Splitting the file by musicianship to carry out further simple effects analyses revealed a much stronger effect in musicians ($F(2,48) = 24.42, p < 0.001, \eta^2 = 0.50$) than in non musicians ($F(2,56) = 3.57, p = 0.04, \eta^2 = 0.11$). Both groups showed a significant upwards trend, but only musicians showed a significant increase in confusion ratings at each level of that trend (from *normal* to *incongruous-resolved*, $p = 0.02$, from *incongruous-resolved* to *incongruous-unresolved*, $p < 0.001$). In non musicians, the only significant difference was an increase in rating from *normal* to *incongruous-unresolved*. Comparing these data to the data from the *oddity* scale, non musicians tend to show higher increases in oddity rating than in confusion rating between each level of *congruence*. This suggests that either the incongruous nature of stimuli does not necessarily incur greater confusion in non musicians (i.e. they knew the stimulus sounded odd but were not troubled when trying to make sense of it and had not been trained to try to maintain the meter) or that this group struggled to apply the notion of confusion to music.

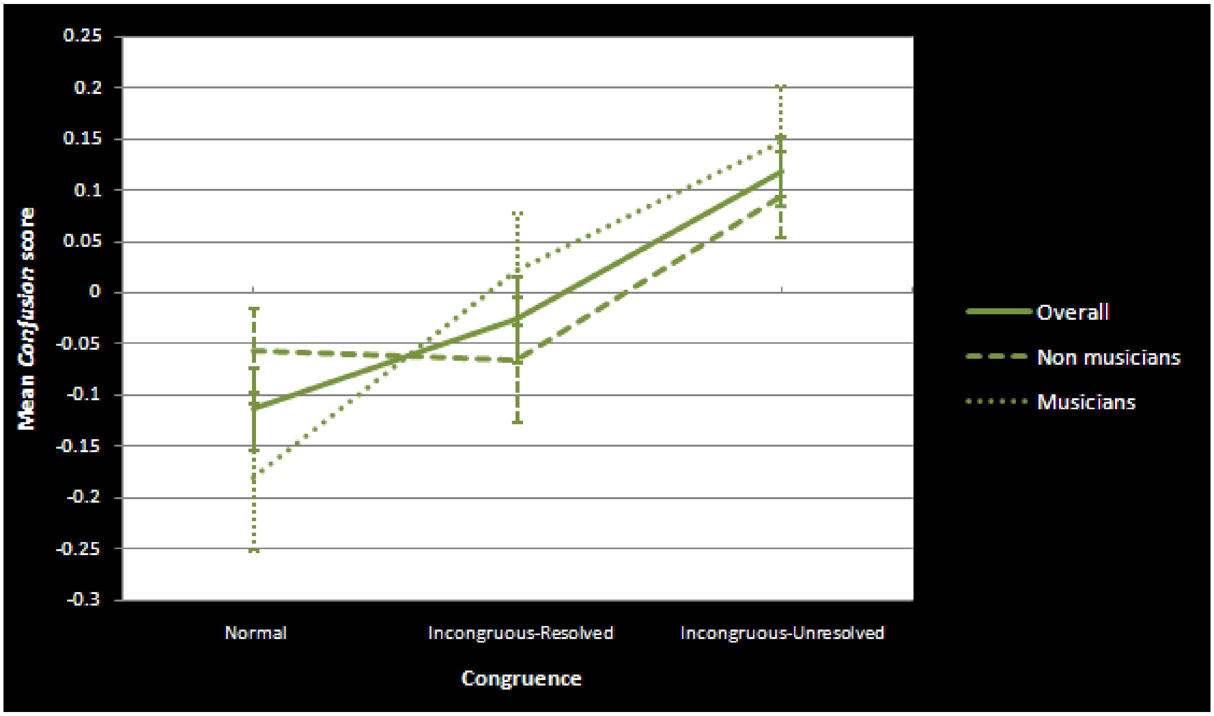

Figure 7.12: *Confusion* ratings by *congruence* within *rhythm* split by *musicianship*

A significant main effect for *congruence* was also seen within *rhythm* ($F(2, 108) = 6.26, p = 0.003, \eta^2 = 0.10$), with the only significant main effect being *normal* stimuli receiving a significantly lower *confusion* rating than *incongruous-unresolved* stimuli ($p = 0.002$). Once more, although no signigicant *congruence* x *musicianship* was found ($F(2, 406) = 1.47, p = 0.26, \eta^2 = 0.03$), the difference in trends observed in Figure 7.12 prompted further simple effects analyses within each level of *musicianship*.

These revealed that the effect of *congruence* was only significant in musicians ($F(2, 48) = 5.01, p = 0.01, \eta^2 = 0.17$, non musicians $F(2, 58) = 2.34, p = 0.11, \eta^2 = 0.08$), where the only significant pairwise comparison was a significantly higher confusion score in *incongruous-unresolved* stimuli than in *normal* stimuli.

The *semantics* stimuli showed a significant main effect for *congruence* ($F(2, 106) =$

$410.93, p < 0.001, \eta^2 = 0.89$), with a significant increase in *confusion* rating at each level of *congruence* (all differences $p < 0.001$), and a significant *congruence* x *musicianship* interaction ($F(2, 104) = 6.30, p = 0.0.003, \eta|^2 = 0.11$) despite the strong overlap in trends between both groups (see Figure 7.13).

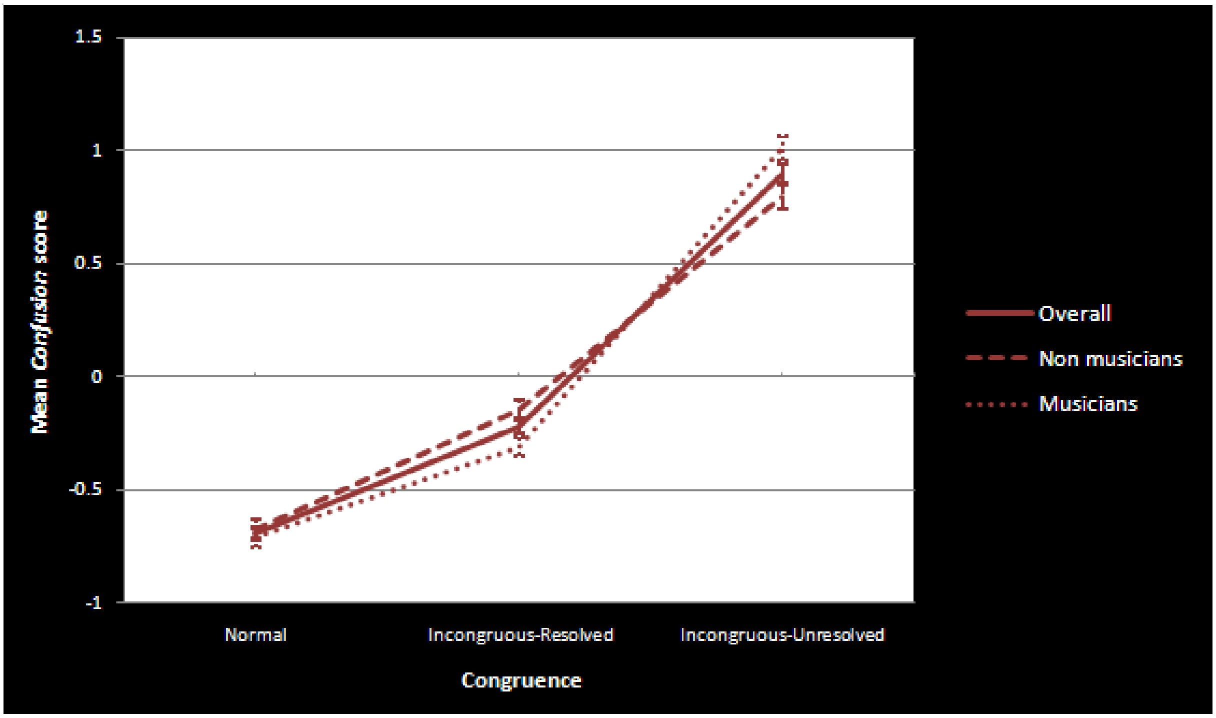

Figure 7.13: *Confusion* ratings by *congruence* within *semantics* split by *musicianship*

Simple effects analyses revealed no differences in significance patterns between groups. The interaction arose from the fact that the difference between *normal* and *incongruous-resolved* stimuli was smaller in musicians (difference between means, 0.40) than in non musicians (difference between means, 0.53) while the difference between *incongruous-resolved* and *incongruous-unresolved* was larger in musicians (difference between means, 1.32) than in non musicians (difference between means, 0.94). Thus, sentences containing resolved incongruities were rated less far along the confusion spectrum in musicians than in non musicians. A percentage value was obtained by dividing the difference between *normal* and *incongruous-resolved* by the difference between *normal* and *incongruous-unresolved* to examine whether any differences were observed between the proportion of the *confusion* scale allocated to *incongruous-resolved* stimuli between *mu-*

sicianship groups. The difference between the mean percentage in musicians (24.94%) and the mean percentage value in non musicians (40.41%) is statistically significant in the one-tailed analysis following this observation ($t(52) = 1.80, p = 0.04$), suggesting that the degree of total confusion experienced in the *incongruous-resolved* condition was significantly smaller in musicians than in non musicians.

Within *syntax*, the same pattern was found: a significant main effect for *congruence* ($F(2, 106) = 241.05, p < 0.001, \eta^2 = 0.82$) with a significant increase at each level of *congruence* (all $p < 0.001$) and no *congruence* x *musicianship* interaction ($F(2, 104) = 0.65, p = 0.52, \eta^2 = 0.01$). The *confusion* ratings showed strong upwards linear trends in the language stimuli in both musicianship groups, although musicians rated *incongruous-resolved semantic* stimuli as less confusing compared to *normal* stimuli than non musicians. In the music stimuli, the linear trend was most evident in results from musicians with data trends similar to those observed in the *oddity* ratings.

7.3.5 Stimulation

The *stimulation* data were participants' responses on the scale "How stimulating did you find the stimulus?" rated from "not at all" to "intensely". The 2 x 3 x 4 ANOVA showed a significant main effect for *congruence* ($F(2, 100) = 7.32, p = 0.002, \eta^2 = 0.13$). Overall, the *normal* stimuli were rated as significantly less stimulating than both *congruence* conditions containing an incongruity (difference with *incongruous-resolved* $p < 0.001$ and with *incongruous-unresolved*, $p = 0.03$), but these two did not differ significantly ($p = 1.00$).

No significant three-way interaction was found ($F(6, 300) = 1.20, p = 0.30, \eta^2 = 0.02$), but both the *hrss* x *congruence* interaction ($F(6, 300) = 2.64, p = 0.03, \eta^2 = 0.05$) and the *hrss* x *musicianship* interaction ($F(3, 150) = 5.01, p = 0.03, \eta^2 = 0.13$) were significant. This last interaction was due to musicians finding the *rhythm* stimuli overall significantly more stimulating than the non musicians did (one tailed $t(52) = 2.28, p = 0.03$), once more possibly because of the increased tension caused by greater efforts in

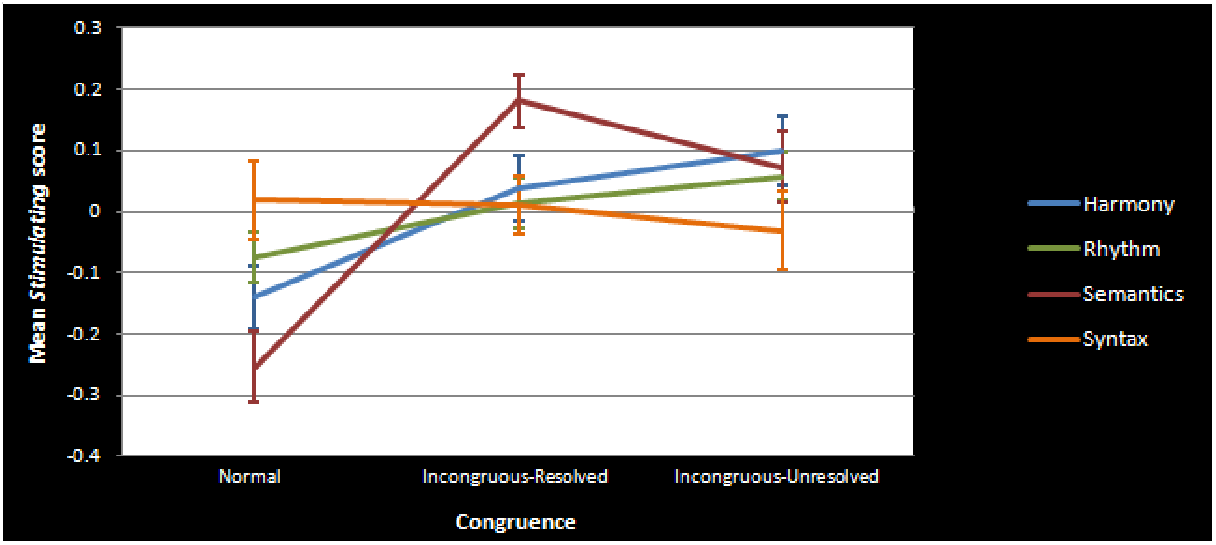

Figure 7.14: *Stimulation* ratings by *hrss* and by *congruence*

this group to maintain the meter. No significant differences between musicians and non musicians were found in the other levels of *hrss*.

Within *harmony*, the simple effects analysis revealed a significant main effect for *congruence* ($F(2, 106) = 3.56, p = 0.03$) with the only significant difference being between *normal* and *incongruous-unresolved*, with the latter receiving a significantly higher rating than the former ($p = 0.03$). Though the *congruence* x *musicianship* interaction was not significant ($F(2, 104) = 1.01, p = 0.37, \eta^2 = 0.02$), the trends shown in Figure 7.15 prompted a further simple effects analyses within each level of *musicianship*.

These revealed that the effect of *congruence* was only significant in musicians ($F(2, 46) = 4.05, p = 0.024, \eta^2 = 0.15$, non musicians $F(2, 58) = 0.64, p = 0.53, \eta^2 = 0.02$). The only significant difference was the *normal* condition receiving a significantly lower stimulation rating than the *incongruous-unresolved* condition.

Within *rhythm*, neither the main effect for *congruence* ($F(2, 104) = 1.76, p = 0.18, \eta^2 = 0.03$) nor the *congruence* x *musicianship* interaction ($F(2, 104) = 1.42, p = 0.25, \eta^2 = 0.03$) were significant. Although the trends displayed in Figure 7.16 suggest different trends in both groups, neither group showed a significant main effect for *congruence*, despite a stronger effect size in non musicians ($\eta^2 = 0.09$) than in non musicians ($\eta^2 = 0.03$). Thus, the manipulation of *rhythm* stimuli did not significantly

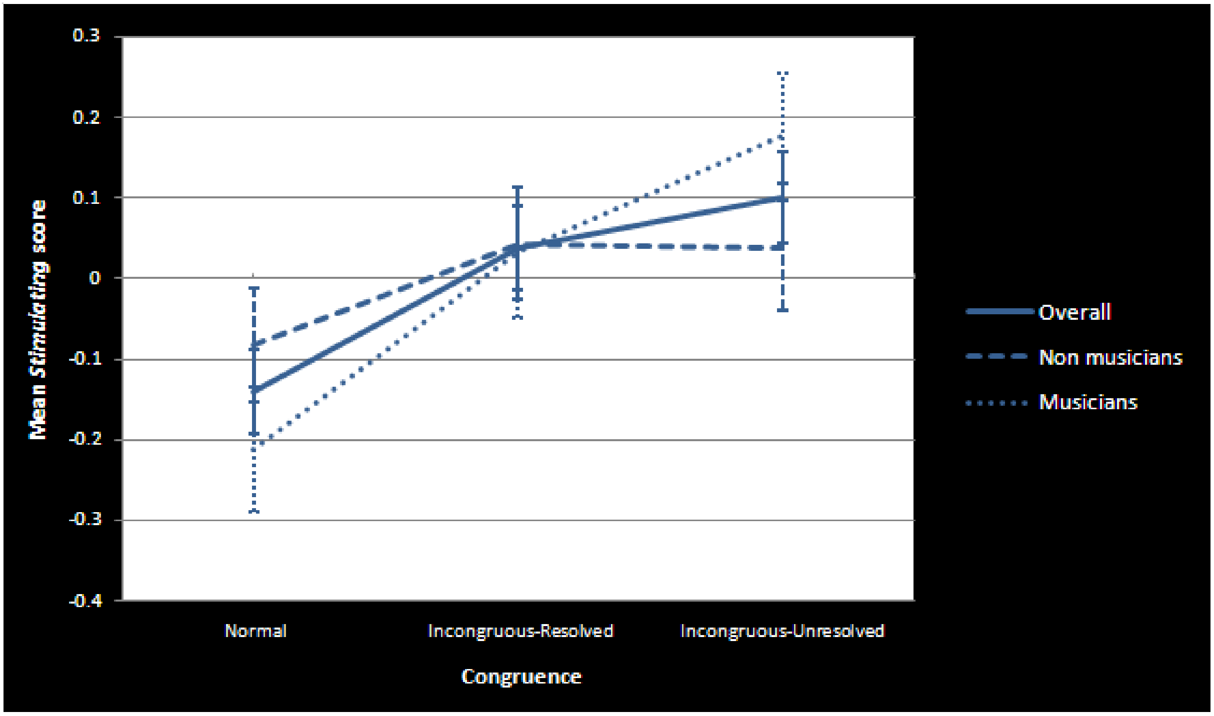

Figure 7.15: *Stimulation* ratings by *congruence* within *harmony* split by *musicianship*

affect how stimulated either musicians or non musicians felt.

As in the case of *harmony*, the *semantics* data, displayed in Figure 7.17, showed a significant main effect for *congruence* ($F(2, 106) = 12.93, p < 0.001, \eta^2 = 0.19$) with the *normal* condition being rated significantly less stimulating than both *incongruous-resolved* ($p < 0.001$) and *incongruous-unresolved* ($p = 0.01$).

No significant *congruence* x *musicianship* effect was found ($\mathrm{F}(2, 104) = 0.31, p = 0.70, \eta^2 = 0.006$), though simple main effects carried out within each level of musicianship showed that the main effect for *congruence* narrowly failed to reach significance in the musician group ($F(2, 48) = 2.93, p = 0.06, \eta^2 = 0.11$, non musicians $F(2, 56) = 11.12, p < 0.001, \eta^2 = 0.28$). In the non-musician group, significant increases in stimulation ratings were observed from *normal* to *incongruous-resolved* ($p < 0.001$) and from *normal* to *incongruous-unresolved* ($p = 0.01$). Thus, contrary to the *harmony* data, this time the non musicians showed a stronger effect of *congruence* than the musicians.

Within *syntax*, no significant effect for *congruence* was found ($F(2, 106) = 0.08, p =$

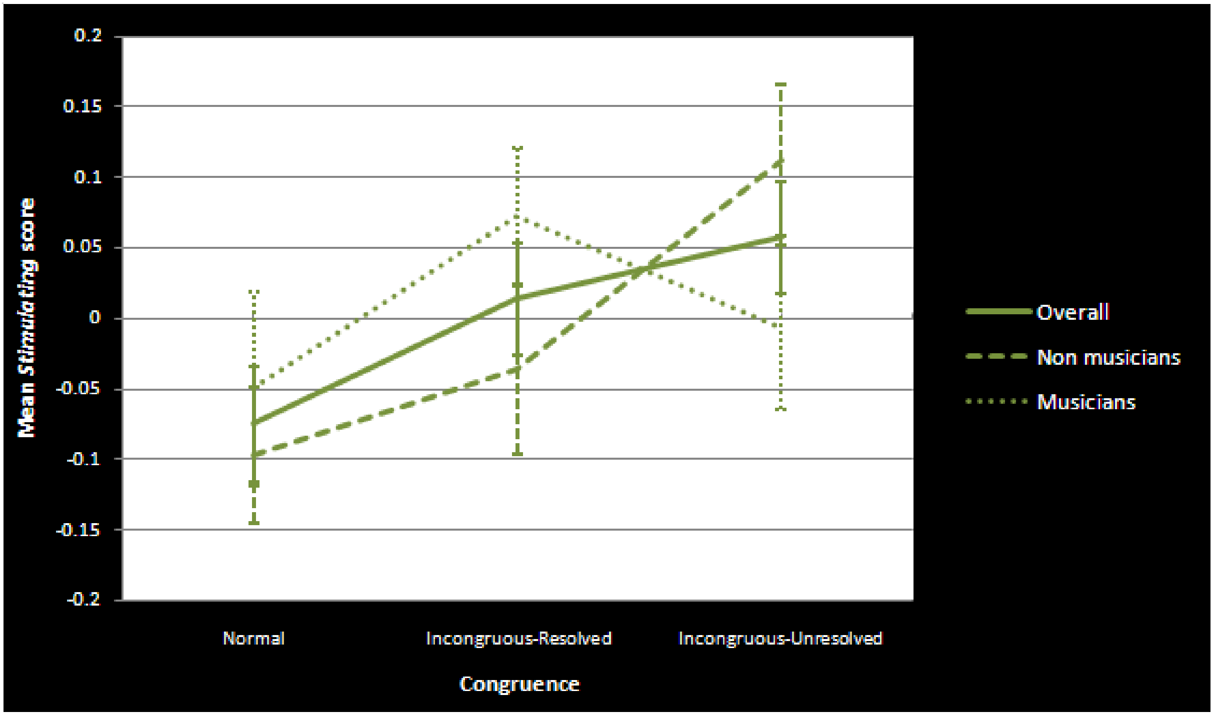

Figure 7.16: *Stimulation* ratings by *congruence* within *rhythm* split by *musicianship*

$0.92, \eta^2 = 0.002$), and the differences between trends (see Figure 7.18) were also non significant ($F(2, 106) = 0.73, p = 0.47, \eta^2 = 0.01$) suggesting that, as in the case of *rhythm*, the experimental manipulation of incongruities in *syntax* had no significant bearing on participants' responses to the stimuli.

The *stimulating* ratings showed mixed results. In *harmony*, the effect was only present in musicians, whereas in *semantics* only non musicians showed a significant main effect for *congruence*. In the other two levels of *hrss*, the main effect of *congruence* failed to reach significance, but the effect was stronger in *rhythm* than in *syntax*. These data contribute to the emerging picture of similar data patterns in *harmony* and *semantic* and to minimal effects being observed in *rhythm* and *syntax*.

7.3.6 Tension during listening

How tense participants felt while listening to the stimulus was assessed after the participants had heard the whole stimulus using the scale "How tense did you feel while listening to the stimulus?" with the anchors "not at all" and "very". The 2 x 3 x 4 ANOVA re-

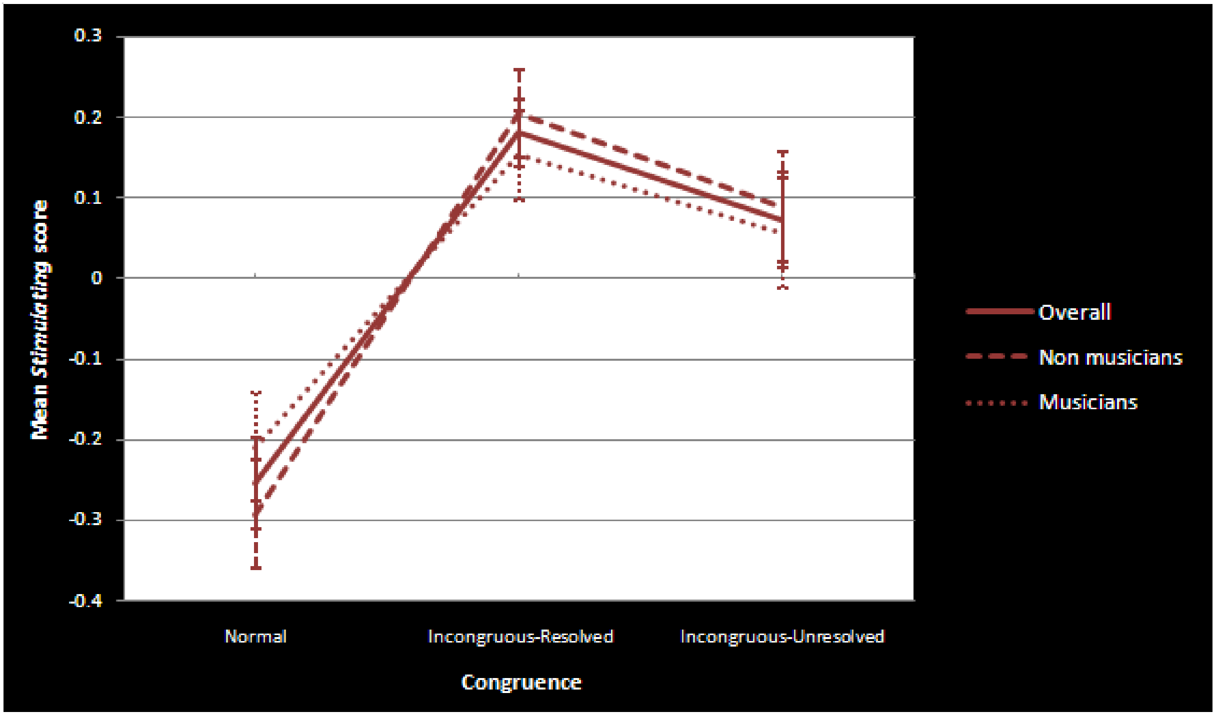

Figure 7.17: *Stimulation* ratings by *congruence* within *semantics* split by *musicianship*

vealed a significant main effect for *congruence* ($F(2, 100) = 54.28, p < 0.001, \eta^2 = 0.52$), a significant *hrss* x *congruence* interaction ($F(6, 300) = 7.40, p < 0.001, \eta^2 = 0.13$) but no significant three-way interaction ($F(6, 300) = 1.15, p = 0.35, \eta^2 = 0.02$). The main effect for *congruence* showed an upwards linear trend with a significant difference at each step, meaning that participants felt more tense while listening to *incongruous-unresolved* stimuli than *incongruous-resolved* stimuli ($p < 0.001$) and more tense while listening to *incongruous-resolved* stimuli than *normal* stimuli ($p < 0.001$).

Within *harmony*, a significant main effect for *congruence* was found ($F(2, 106) = 7.22, p = 0.001, \eta^2 = 0.12$) following which pairwise comparisons revealed that the *incongruous-unresolved* condition resulted in significantly higher tension ratings than both the *normal* ($p = 0.002$) and *incongruous-resolved* ($p = 0.048$) conditions.

Though the *congruence* x *musicianship* interaction suggested by the data shown in Figure 7.20 was not significant ($F(2, 104) = 2.06, p = 0.13, \eta^2 = 0.04$), further simple effects analyses revealed that the effect of *congruence* was only significant in musicians ($F(2, 46) = 7.53, p = 0.001, \eta^2 = 0.25$, non musicians $F(2.27), p = 0.14, \eta^2 = 0.07$). In

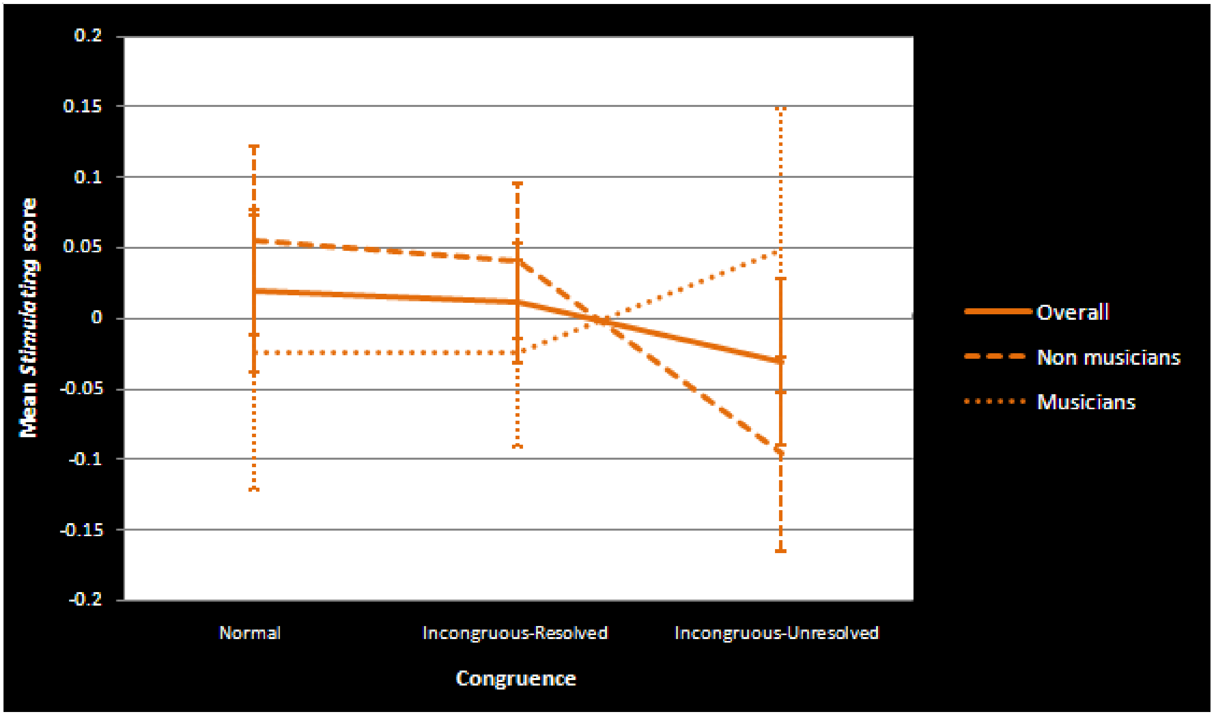

Figure 7.18: *Stimulation* ratings by *congruence* within *syntax* split by *musicianship*

musicians, the same significance patterns were observed as when all participants were pooled together, namely a significantly higher tension rating in *incongruous-unresolved* than in *incongruous-resolved* ($p = 0.006$) and than in *normal* ($p = 0.014$).

Within *rhythm* also, a significant main effect for *congruence* was seen ($F(2, 106) = 7.29, p = 0.001, \eta^2 = 0.12$). This time, the significant difference lay between the *normal* condition receiving a significantly lower tension rating than both the *incongruous-resolved* condition ($p = 0.007$) and the *incongruous-unresolved* condition ($p < 0.001$). Once again, the interaction suggested by the data shown in Figure 7.21 failed to reach significance ($F(2, 104) = 2.45, p = 0.09, \eta^2 = 0.045$).

However, simple effects analyses within each level of *musicianship* showed that the effect of *congruence* on tension ratings was only significant in musicians ($F(2, 48) = 10.08, p < 0.001, \eta^2 = 0.30$, non musicians $F(2, 56) = 1.00, p = 0.38, \eta^2 = 0.03$) who showed the same significance pattern as the overall data (*normal* significantly lower than *incongruous-resolved* ($p = 0.002$) and than *incongruous-unresolved* ($p < 0.001$). Thus, in both *harmony* and *rhythm*, tension ratings were only significantly affected by

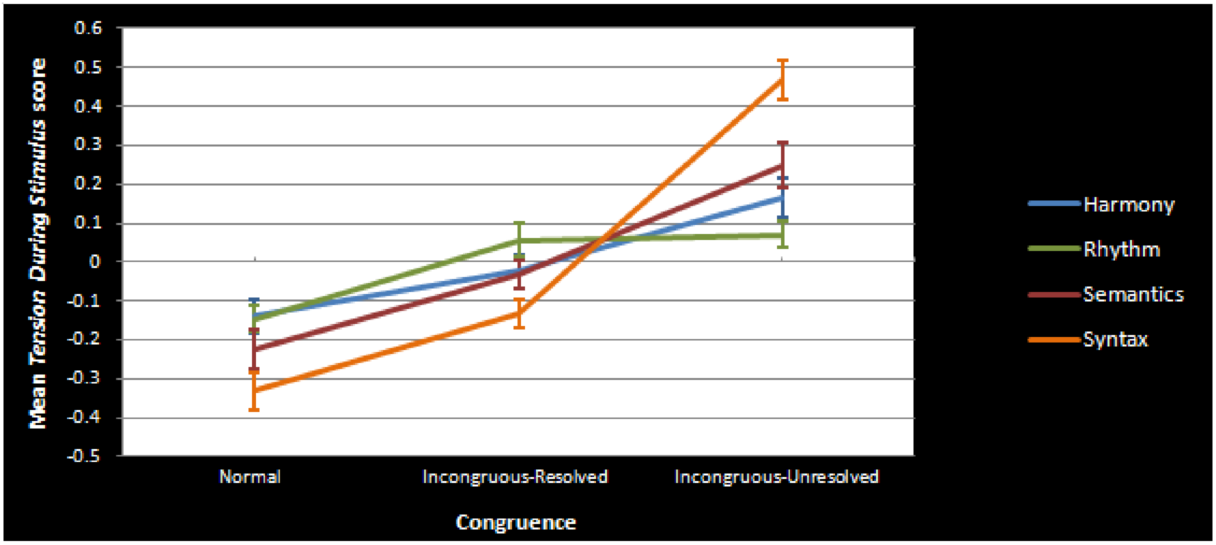

Figure 7.19: *Tension During Listening* ratings by *hrss* and by *congruence*

the stimulus manipulation in musicians.

The *semantics* stimuli also showed a significant main effect for *congruence* ($F(2, 106) = 16.13, p < 0.001, \eta^2 = 0.23$), with a significant increase in tension ratings between all levels of *congruence* in an upwards linear trend (*normal* < *incongruous-resolved* < *incongruous-unresolved*, differences respectively $p = 0.015$ and $p = 0.003$). Though the *congruence* x *musicianship* interaction suggested by Figure 7.22 was not significant ($F(2, 104) = 0.24, p = 0.74, \eta^2 = 0.005$), splitting the data by *musicianship* revealed different patterns between groups.

Splitting the file by musicianship revealed that the effect of *congruence* in musicians ($F(2, 48) = 9.46, p = 0.001, \eta^2 = 0.28$) resulted in *incongruous-unresolved* stimuli receiving a significantly higher rating than the other two levels of *congruence* (both differences $p = 0.004$), with no significant difference between *incongruous-resolved* and *incongruous-unresolved*, whereas in the non musicians, the main effect ($F(2, 56) = 6.70, p = 0.005, \eta^2 = 0.19$) resulted in the *incongruous-resolved* stimuli receiving a significantly higher rating than the other two stimuli (both differences $p < 0.05$) with no significant difference between *normal* and *incongruous-resolved.* Thus, only the musicians rated the *incongruous-resolved* stimuli as significantly more tension-inducing than the *normal* stimuli.

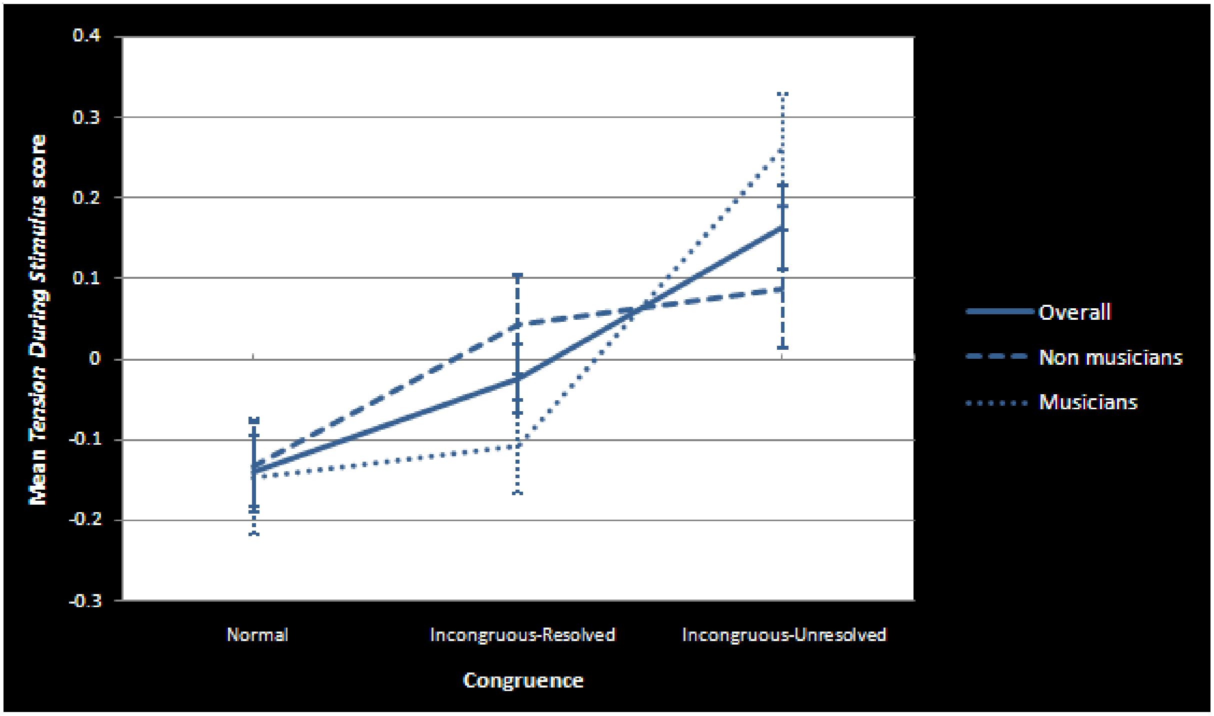

Figure 7.20: *Tension During Listening* ratings by *congruence* within *harmony* split by *musicianship*

Finally, in *syntax* a strong significant main effect for *congruence* was seen ($F(2, 108) = 55.63, p < 0.001, \eta^2 = 0.51$) with a significant increase in tension ratings from *normal* to *incongruous-resolved* ($p = 0.001$) and from *incongruous-resolved* to *incongruous-unresolved* ($p < 0.001$). Though the *congruence* x *musicianship* interaction failed to reach significance ($F(2, 106) = 2.35, p = 0.11, \eta^2 = 0.04$), simple effect analyses within each level of musicianship revealed that once again only in the musician group was the *incongruous-resolved* condition rated significantly more tension-inducing than the *normal* condition.

The *tension during listening* ratings overall showed a gradual increase from *normal* to *incongruous-resolved* and to *incongruous-unresolved* in all levels of *hrss*. Differences between musicians and non musicians lay in the strength of the effects and in the significance of differences between the *incongruous-resolved* condition and the other two *congruence* conditions.

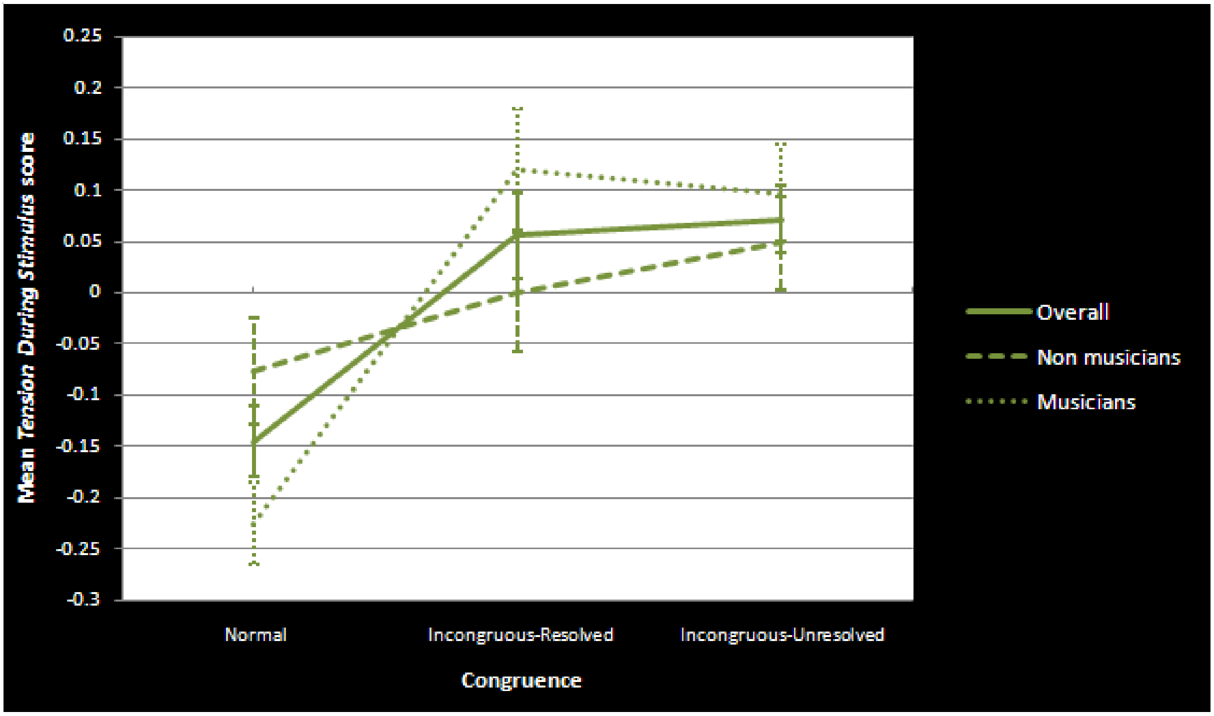

Figure 7.21: *Tension During Listening* ratings by *congruence* within *rhythm* split by *musicianship*

7.3.7 Tension after listening

Finally, the ratings of how tense participants felt after listening to the stimulus, collected using the scale "How do you feel now, after having listened to the whole stimulus" ranging from "very relaxed" to "very tense", showed a significant main effect for *congruence* ($F(2, 100) = 56.10, p < 0.001, \eta^2 = 0.53$) and a significant interaction of *hrss* x *congruence* ($F(6, 300) = 10.34, p < 0.001, eta^2 = 0.17$), but no significant three-way interaction with *musicianship* ($F(6, 300) = 0.92, p = 0.46, \eta^2 = 0.02$).

Pairwise comparisons revealed a significant increase in tension ratings from *normal* to *incongruous-resolved* ($p = 0.05$) and again from *incongruous-resolved* to *incongruous-unresolved* ($p < 0.001$).

Within *harmony* a significant main effect was found for *congruence* ($F(2, 106) = 14.92, p < 0.001, \eta^2 = 0.22$) with pairwise comparisons revealing that both *normal* and *incongruous-resolved* received a significantly lower tension rating (indicating a greater degree of relaxation) than *incongruous-unresolved* stimuli ($p < 0.001$). No significant

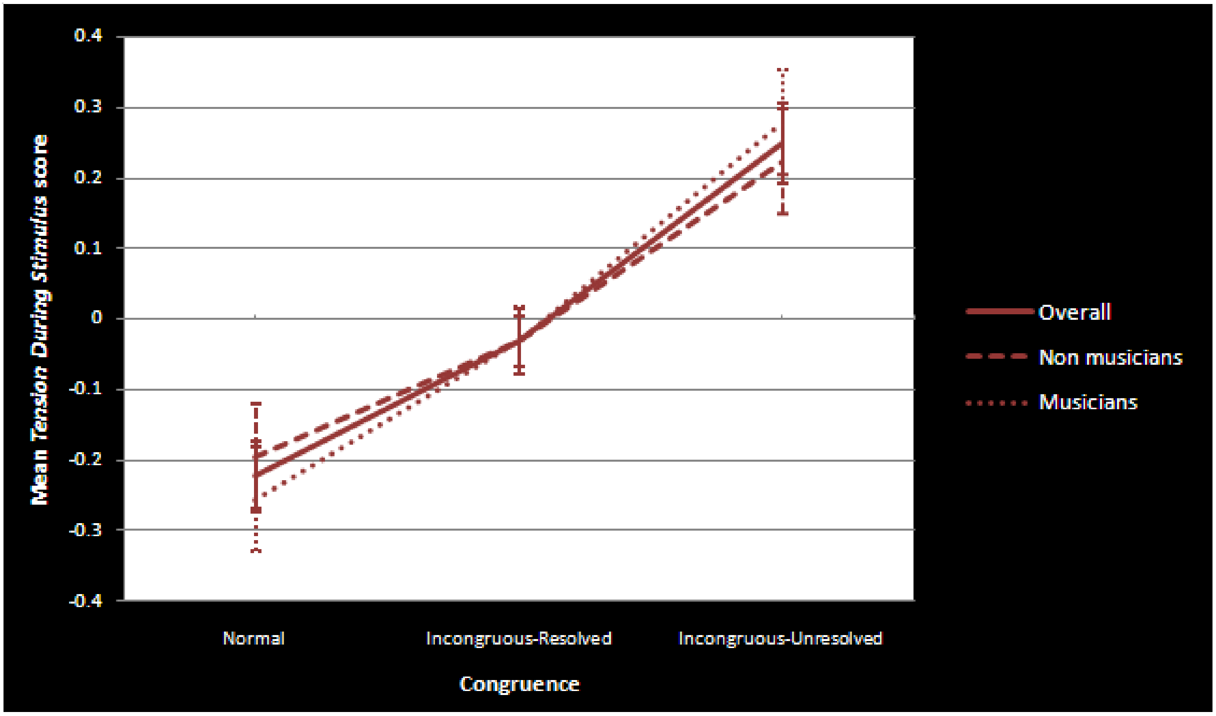

Figure 7.22: *Tension During Listening* ratings by *congruence* within *semantics* split by *musicianship*

difference was seen between *normal* and *incongruous-resolved* stimuli. No significant *congruence* x *musicianship* interaction was found ($F(2, 106) = 0.94, p = 0.39, \eta^2 = 0.02$), despite the differences in the trends shown in Figure 7.24 in which a slight decrease from *normal* to *incongruous-resolved* was seen in musicians while a slight increase from *normal* to *incongruous-resolved* was seen in non musicians.

The main effect for *congruence* was strongest in the non-musician group ($F(2, 58) = 16.92, p < 0.001, \eta^2 = 0.37$) where pairwise comparisons showed a significantly higher rating in *incongruous-unresolved* compared to both *normal* ($p < 0.001$) and *incongruous-resolved* ($p = 0.002$). In the musician group ($F(2, 48) = 3.79, p = 0.03, \eta^2 = 0.14$) the only significant difference was a higher rating in *incongruous-unresolved* than in *incongruous-resolved*, with a slight but non significant decrease between *normal* and *incongruous-resolved*. In line with this observation, a linear trend was significant ($p < 0.001$) only in non musicians.

Within *rhythm*, both the main effect for *congruence* ($F(2, 106) = 0.08, p = 0.921, \eta^2 = 0.002$) and the *congruence* x *musicianship* interaction ($F(2, 106) = 0.40, p = 0.67, \eta^2 =$

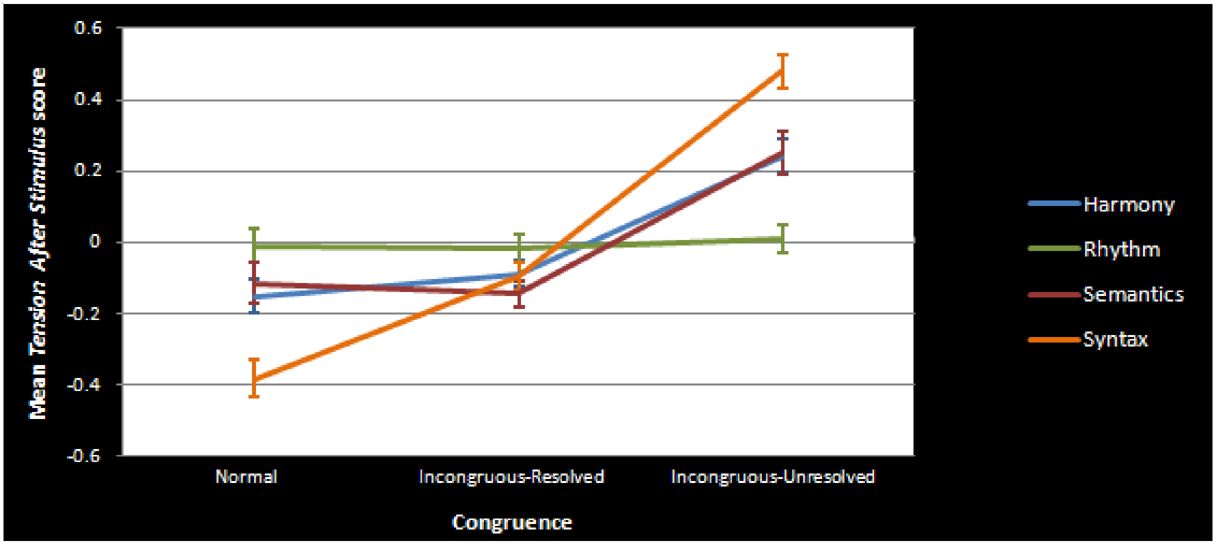

Figure 7.23: *Tension After Listening* ratings by *hrss* and by *congruence*

0.007) were far from significant. These data indicate that the experimental manipulation in stimuli had very little effect on how tense participants felt after listening to the stimulus in both groups of participants. The fact that a significant effect was found in ratings of how tense participants felt *while* they were listening to the stimulus suggests that participants were able to distinguish between these two scales.

Congruence had a significant effect within *semantics* ($F(2, 104) = 11.81, p < 0.001, \eta^2 = 0.19$), with the *incongruous-unresolved* condition receiving a significantly higher tension rating than both *normal* ($p = 0.005$) and *incongruous-resolved* ($p < 0.001$). Once more, the *congruence* x *musicianship* interaction suggested in Figure 7.25 was not significant ($F(2, 102) = 0.76, p = 044, \eta^2 = 0.02$).

Contrary to the *harmony* data, this time the effect of *congruence* was strongest in non musicians ($F(2, 54) = 7.25, p = 0.004, \eta^2 = 0.21$, musicians $F(2, 48) = 5.37, p = 0.01, \eta^2 = 0.18$), and non musicians showed a strong quadratic trend ($F(1, 27) = 15.54, p = 0.001, \eta^2 = 0.37$) while the strongest trend in musicians was linear ($F(1, 24) = 6.43, p = 0.02, \eta^2 = 0.21$). In both groups, the only significant pairwise comparison was the *incongruous-unresolved* condition receiving a significantly higher rating than the *incongruous-resolved* condition. However, while in *harmony* the musicians showed a slight decrease from *normal* to *incongruous-resolved* and the non musicians showed a

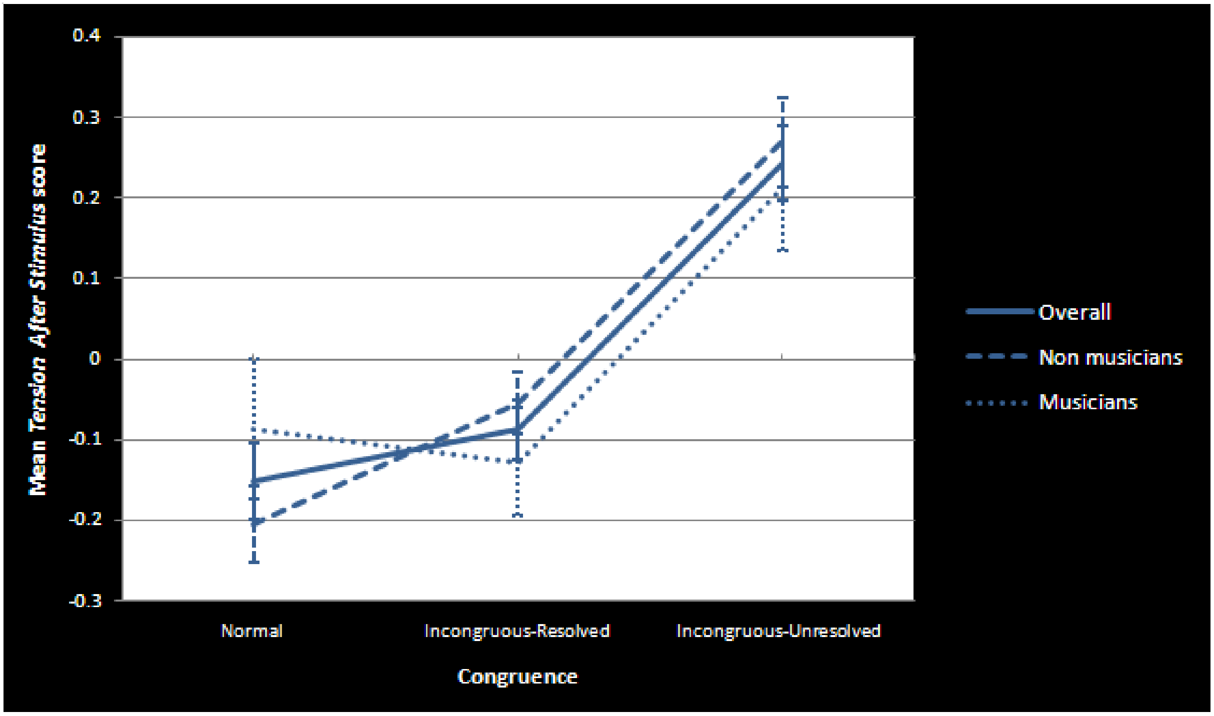

Figure 7.24: *Tension After Listening* ratings by *congruence* within *harmony* split by *musicianship*

slight increase between those two conditions, in *semantics*, it is the non musicians who showed the slight decrease and the musicians who showed the slight increase.

Within *syntax*, a significant main effect for *congruence* was observed ($F(2, 104) = 60.18, p < 0.001, \eta^2 = 0.54$), with a significant increase in tension ratings at each level of the linear trend (*normal* to *incongruous-resolved*, $p = 0.001$) and *incongruous-resolved* to *incongruous-unresolved*, $p < 0.001$). The trends were very similar across participant groups, and no significant *congruence* x *musicianship* interaction was found ($F(2, 104) = 1.13, p = 0.33, \eta^2 = 0.02$).

The *tension after the stimulus* ratings showed a significant main effect for *congruence* in *harmony*, *semantics* and *syntax* with a reversal of patterns between musicians and non musicians in these first two levels of *hrss* as in the *aesthetic* ratings. Similarities such as these showing how the two groups compare to each other across different ratings are the subject of the next section of this chapter.

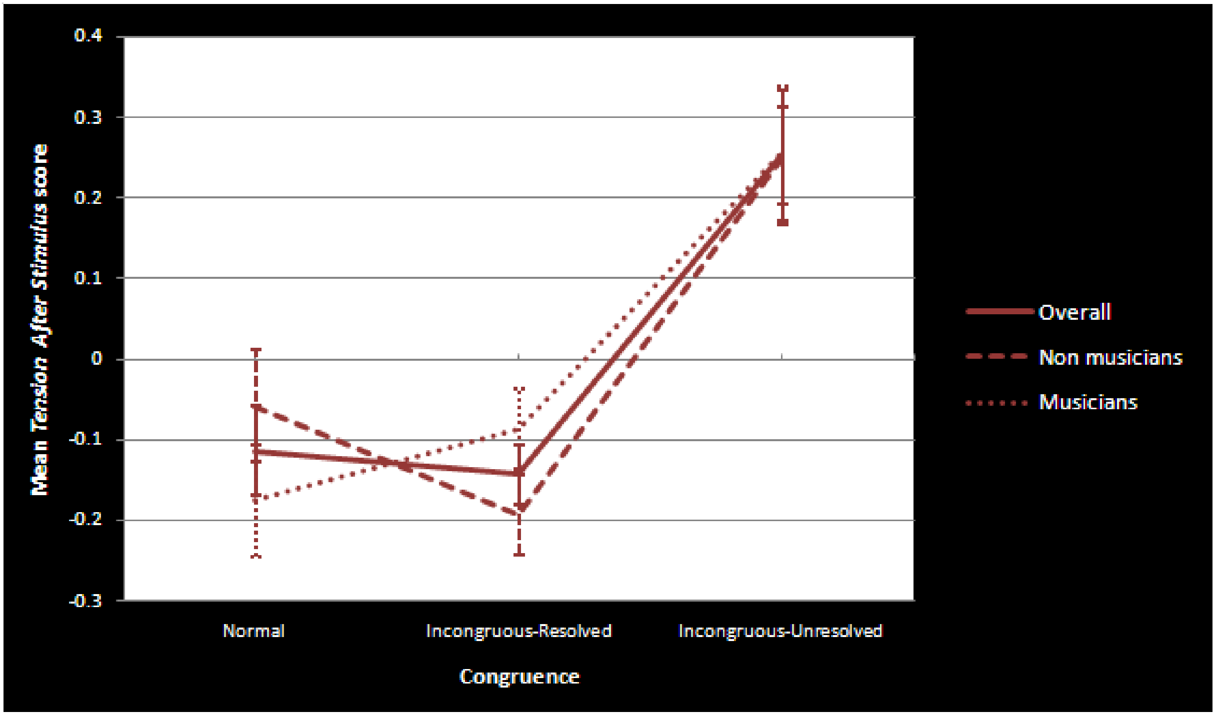

Figure 7.25: *Tension After Listening* ratings by *congruence* within *semantics* split by *musicianship*

7.4 Relationships between the rating scales

The aim in collecting data from a variety of scales was to ascertain whether any predictive relationship could be established for the different factors underlying aesthetic appreciation of music and language. Since the *harmony* and the *semantics* stimuli were the only ones to show significant effects of *congruence* on all scales, these alone will form the object of this section.

The summary of the ANOVA outcomes showed similar patterns across *hrss* in terms of which group showed a significant effect of *congruence* on aesthetic ratings, with an apparent cross-over between the trends observed in musicians and non musicians, with similar patterns happening in harmony in non musicians and in semantics in musicians. These effects, summarised in Table 7.2 were most noticeable in the ratings relating to stimulation and tension.

7.4.1 Regression analysis

In order to assess the statistical reliability of these observations, a multiple regression analysis was carried, using *aesthetic pleasantness* as the predicted variable and all seven other scales as predictors, in a stepwise model. To make the scales comparable for the purpose of this analysis, z-scores were calculated within each level of *hrss* for each scale across all participants.

The regression was carried out separately for both levels of *hrss*, to compare the significance of different predictors across domains. As shown in Table 7.1, in both *harmony* and *semantics*, the two strongest predictors retained for the final model were, in order of contribution, *Tension after listening* and *Stimulation*. In *harmony*, the third predictor was *Tension during listening* whereas in *semantics* that predictor was *Confusion*.

Table 7.1: Multiple regression outcome for *harmony* and *semantics* with the significant predictors of *aesthetic* ratings by *hrss*

Harmony	Model	F(3,116)=14.64, p<0.001		
	Coefficients	Standardased coefficient	t-value	Significance
	Tension After	β = -0.47	t = -5.44	p < 0.001
	Stimulation	β = 0.31	t = 3.75	p < 0.001
	Tension During	β = 0.2	t = 2.22	p = 0.03
Semantics	Model	F(3,116)=61.31, p<0.001		
	Coefficients	Standardased coefficient	t-value	Significance
	Tension After	β = -0.62	t = -9.09	p < 0.001
	Stimulation	β = 0.23	t = 3.82	p < 0.001
	Confusion	β = -0.26	t = -3.73	p < 0.001

This raises the question of whether oddity, via the means of tension and stimulation could underlie aesthetic experiences in both music and language. To examine this, two mediation analyses were envisaged, the first looking at the potential mediating effect of tension between oddity and aesthetics, and the second looking at the potential mediating effect of stimulation between oddity and aesthetics. In both these mediations, a significant correlation would be necessary between oddity ratings and aesthetic ratings. However, although this correlation was significant in *semantics* ($r = -.41, p < 0.001$), it was not significant in *harmony* ($r = -.04, p = 0.65$), rendering this analysis meaningless as an explanation for the role of oddity in aesthetics across both domains.

In *semantics*, the correlation between oddity and aesthetic ratings was significant ($r = -.41, p < 0.001$), but aesthetic ratings were not significantly correlated with stimulation ratings ($r = .14, p = 0.13$), ruling stimulation out as a mediator between oddity and aesthetics. The correlation between tension after listening and aesthetic ratings was significant ($r = -.70, p < 0.001$), as was the correlation between tension after listening and oddity ratings ($r = -0.42, p < 0.001$), but partialling out the effect of tension after listening led to a reduction in the correlation effect (from $r = -.50, p < 0.001$ to $r = -.24, p = 0.01$), thus reducing rather than enhancing the explanatory power of oddity ratings on aesthetic ratings.

7.4.2 Aesthetics or affect?

Throughout this thesis, the term aesthetic-affective has been used as an umbrella term for the responses described in previous publications investing incongruity processing in an aesthetics framework. Up to this point in the analysis, no distinction has been made between aesthetic responses and affective responses. Since it is possible to have one without the other — to notice the beauty of a stimulus without enjoying it, or to experience delight at a stimulus in which one can find no objective beauty — it was important these two facets of "aesthetic delight" (Aiken, 1951, p. 29) were tested separately in this experiment.

The data gathered in Experiment 3 showed a strong and significant negative correlation between *aesthetic pleasantness* ratings and *tension after listening* ratings in all levels of *hrss* (*harmony* $r = -.35, N = 120, p < 0.001$; *rhythm* $r = -.44, N = 120, p < 0.001$; *semantics* $r = -.70, N = 117, p < 0.001$; *syntax* $r = -.83, N = 120, p < 0.001$). The question of whether the greater relaxation followed the perception of aesthetically pleasing stimuli or whether stimuli were perceived as aesthetically pleasing because the participant felt more relaxed after listening to them cannot be answered through this analysis. Nonetheless, the strength of the correlations show that the more aesthetically pleasing stimuli were said to be, the more relaxed participants reported feeling.

7.5 Discussion

The data reported in this chapter first of all provided a validation of the stimulus set, demonstrating that *normal* stimuli were rated as significantly less odd than *incongruous-resolved* stimuli, which in turn were rated as significantly less odd than *incongruous-unresolved* stimuli. The trend was however not as strong in the *rhythm* stimuli, with ratings of oddity being close in the *incongruous-resolved* and the *incongruous-unresolved* condition. As suggested from the data in Chapter 4, this could be due to participants perceiving both the onset and the offset of the different rhythmic pattern as incongruous. The ERP data explored in Chapter 8 seem to support this interpretation. The caution expressed surrounding the inability to determine which meter participants were keeping while listening to the rhythmically ambiguous incongruous sections of the rhythm stimuli could also explain these weaker effects Some participants may have been determined to hang onto the original meter while others may have switched meter and thus not needed further resolution and would thus have experienced the *rhythm* stimuli very differently.

The data also provided support for the role of incongruities in eliciting affective-aesthetic responses to music and language only when the incongruity was followed by a resolution in *harmony* and *semantics*. These findings are the first of their kind in an experiment with such tightly controlled and ecologically valid stimuli in which the experimental manipulation was matched across music and language. The differences between musicianship groups were surprising, notably in the *semantics* stimuli, and the other scales were explored with the aim of providing a statistical model of the relationship between oddity and aesthetics.

This exploration led to the observation of similar patterns in ratings between the non-musician data in *harmony* and the musician data in *semantics*, and between the musician data in *harmony* and the non musicians data in *semantics*. In addressing whether aesthetic ratings could be predicted from the data on the other scales used in this experiment, a regression analysis was carried out. This revealed that in both *harmony* and *semantics*, the strongest predictors of aesthetic ratings were *Stimulation* and *Tension After Listening*.

However, these data do not provide a reason for why different levels of theoretical incongruity should have different effects on stimulation, tension and aesthetic ratings in musicians and non musicians. The exploration of ERP data presented in Chapter 8 aimed to uncover any systematic differences between musicianship groups, in an attempt to understand the role of the brain's response to incongruities in eliciting affective responses to music and language.

Table 7.2: Summary of effects of *congruence* in *harmony* and *semantics*

		Musicians		Non-musicians	
	Scale	Normal to Incong.-R.	Incong.-R. to Incong.-U.	Normal to Incong.-R.	Incong.-R. to Incong.-U.
Harmony	Aesthetic pleasantness	-	-	sig. increase	-
	Oddity	sig. increase	sig. increase	near sig. Increase	sig. increase
	Interest	near sig. Increase	-	-	-
	Confusion	sig. increase	sig. increase	-	-
	Stimulation	-	(sig. increase from N to IU)	-	-
	Tension During Listening	-	sig. increase	large (n.s.) increase	small (n.s.) increase
	Tension After Listening	small (n.s.) decrease	sig. increase (quadratic trend)	small (n.s.) increase	sig. increase (linear trend)
Semantics	Aesthetic pleasantness	sig. increase	-	-	sig. decrease
	Oddity	sig. increase	sig. increase	sig. increase	sig. increase
	Interest	sig. increase		sig. increase	
	Confusion	smaller (sig.) increase	sig. increase	larger (sig.) increase	sig. increase
	Stimulation	-	-	sig. increase	-
	Tension During Listening	sig. increase	-	-	sig. increase
	Tension After Listening	small (n.s.) increase	sig. increase (linear trend)	small (n.s.) decrease	sig. increase (quadratic trend)

Chapter 8

Integration difficulty and blissful ignorance: ERP data from the processing of aesthetically-purposed incongruities and their relationship to behavioural outcomes

8.1 Introduction

This chapter presents the ERP data gathered in Experiment 3, adding the neurophysiological patterns associated with the incongruous element itself to the investigation of the role of incongruities in eliciting aesthetic-affective responses to music and language. The first section provides an exploratory analysis of the ERP effects associated with incongruity processing in all four components of music and language. Where different patterns were observed in musicians and non musicians these were followed up with separate analyses within each group, according to recommendations by Luck (2006). The second section explores ERP effects at the point of resolution providing a means of testing the suggestion that different patterns in harmony ERP research can be accounted for by differences in stimulus design, as well as a means of exploring the reasons underlying the unexpected trends in *rhythm* oddity ratings. The third section aims to provide further insights into the interpretation of ERP effects, by analysing the relationships between the ERP data and the data from the rating scales detailed in Chapter 7. This final section also explores the possibility of ERP effects accounting for the differences between musicians and non musicians in their responses to aesthetically-purposed incongruities in music and language. Once again, results are discussed along with the analyses.

8.2 EEG data preparation

Following recommendations from Luck (2006) for exploratory analyses, the data from the 64 electrodes were averaged into nine clusters, or *regions of interest*. The nine clusters here separated the scalp into three from left to right and into three from front to back. The resulting clusters were left frontal, left central, left parietal, midline frontal, midline central, midline parietal, and right frontal, right central, right parietal. The electrodes included in each cluster were as follows:

left frontal FP1, AF3, F3, F5, F7

midline frontal FPz, F1, Fz, F2

right frontal FP2, AF4, F4, F6, F8

left central FT7, FC5, FC3, C3, C5, T7

midline central FC1, FCz, FC2, C2, Cz, C1

right central FC4, FC6, FT8, C4, C6, T8

left parietal TP7, CP5, CP3, P7, P5, P3

midline parietal CP1, CPz, CP2, P1, Pz, P2

right parietal CP4, CP6, TP8, P4, P6, P8

ERP data were analysed using a 3 (*congruence*) x 3 (anterior-posterior: frontal, central, parietal) x 3 (laterality: left, midline, right) x 2 (musicianship) ANOVA. Significant interactions were followed up with simple effects analyses. A Greenhouse-Geisser correction was used where appropriate. Time-windows for statistical analysis were chosen based on visual inspection of the data and taking into consideration previous research in incongruity processing in music and language.

Since ERP effects are only meaningful by comparison to a baseline, the only meaningful comparisons between the three *congruence*conditions at the point of incongruity were 1) between the mean responses to *normal* stimuli and *incongruous-resolved* stimuli and 2) between the mean responses to *normal* stimuli and *incongruous-unresolved* stimuli. The reported amplitudes are the mean amplitude of the EEG wave over the time-window specified in each analysis.

Simple two-tailed planned contrasts, comparing the two incongruous levels of *congruence* to the *normal* condition were therefore carried out, to determine whether the incongruity led to a significant departure from the pattern observed in the *normal* condition. Where results are compared across musicianship groups, the topographical map for both groups is produced in the same scale. The electrode from which the waveform is taken is highlighted in each topographical map.

8.3 ERP effects at the point of incongruity

8.3.1 Harmony

When comparing waveforms for incongruous vs normal stimuli, the two waveforms showed a good pre-stimulus overlap and a divergence from 400 to 800 ms, as seen in Figure 8.1. Topographical maps showed that the time-window which best captures this effect in terms of its distribution on the scalp was from 500 to 700ms, the same time-window as the one explored in Patel et al.'s (1998) study of harmonic and syntactic processing. This effect seemed to result in a difference between *congruence* conditions in the central and posterior regions of the scalp. When looking at the effects within each participant group, musicians seemed to show a stronger effect than non musicians, and this effect seemed to be slightly more right lateralised.

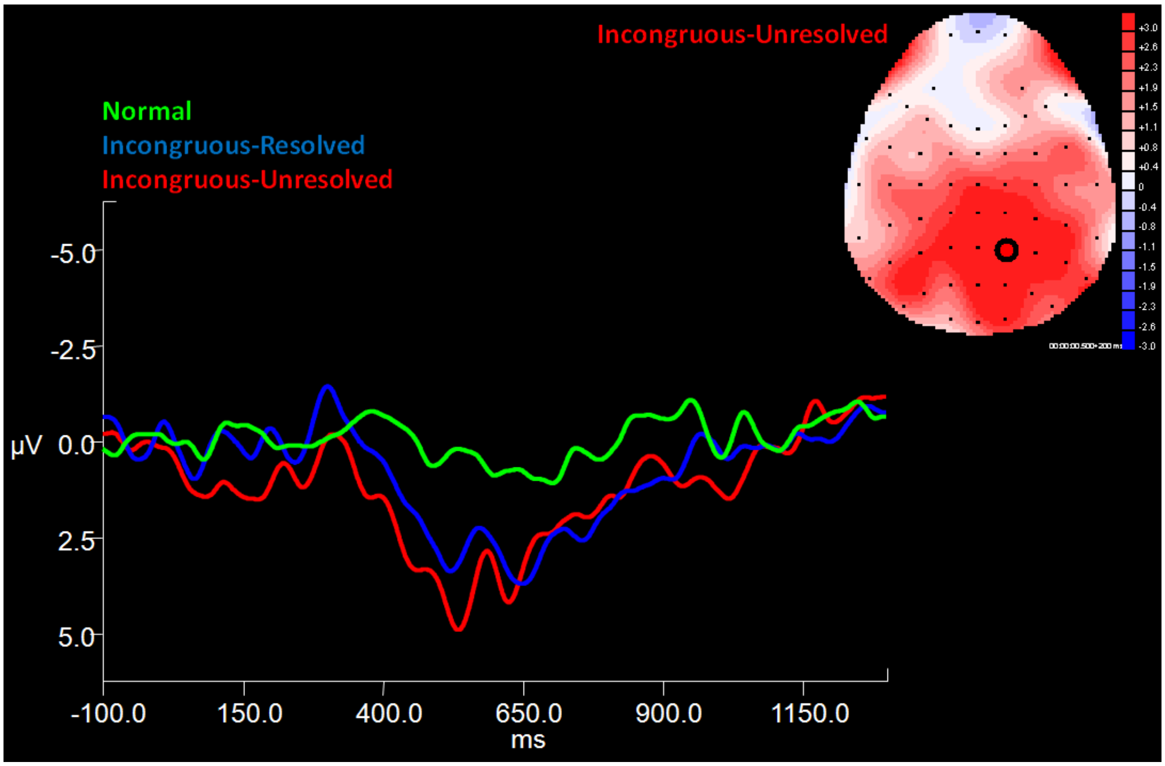

Figure 8.1: Waveforms for all three levels of *congruence* and topographical map for *incongruous-unresolved* minus *normal* in the 500 to 700ms time-window after the incongruity in *harmony* in *musicians*

8.3.1.1 Results

The initial overall ANOVA showed no significant main effect for *congruence* ($F < 1$), but a significant *congruence* x *anterior-posterior* interaction ($F(4, 152) = 4.91, p < 0.01, \eta^2 = 0.11$). Simple effects analyses revealed that the effect for *congruence* was significant throughout the parietal clusters ($F(2, 78) = 3.15, p = 0.05, \eta^2 = 0.07$), where no significant interaction effect of *congruence* with *laterality* was found. Planned contrasts across all three parietal clusters revealed a significant relative positivity only in the *incongruous-unresolved* condition compared to the *normal* condition ($F(1, 39) =$

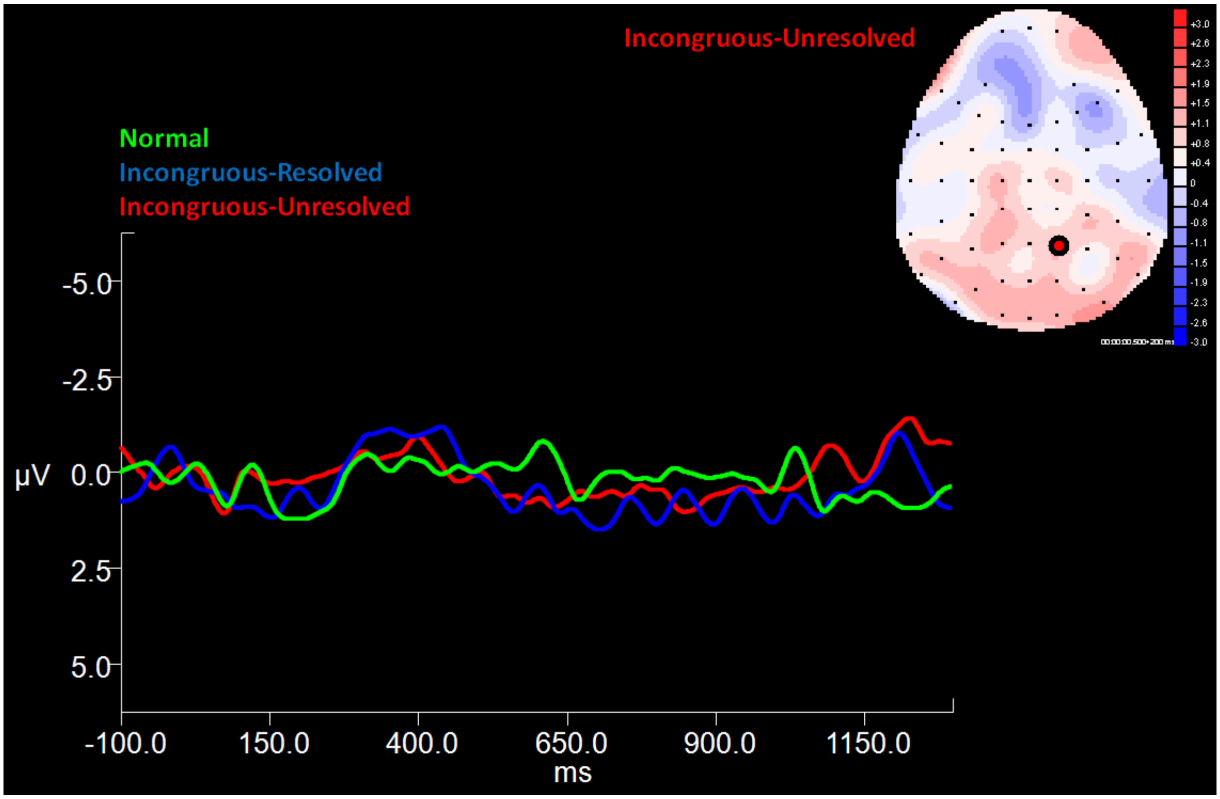

Figure 8.2: Waveforms for all three levels of *congruence* and topographical map for *incongruous-unresolved* minus *normal* in *harmony* in the 500 to 700ms time-window after the incongruity in *non musicians*

$5.11, p = 0.03, \eta^2 = 0.12$), but no significant ERP effect in the *incongruous-resolved* condition in comparison to the *normal* condition.

Though no significant *musicianship* x *congruence* effect had been observed, the data were split by musicianship in view of the differences between groups displayed in Figures 8.1 and 8.2. As suggested by Figure 8.2, non musicians showed no significant main effect for *congruence* and no significant interactions of *congruence* with any of the localisation factors, meaning that incongruous stimuli did not result in a significantly different amplitude to normal stimuli at 500 to 700ms after the point of incongruity in non musicians.

Musicians on the other hand showed no significant main effect for *congruence* but a significant *congruence* x *anterior-posterior* interaction effect ($F(4, 76) = 5.57, p < 0.01, \eta^2 = 0.23$). Once again, this interaction was followed up by simple effect analyses within each level of anterior-posterior, revealing a significant main effect for *con-*

gruence in the parietal clusters ($F(2,38) = 4.73, p = 0.02, \eta^s = 0.20$) where a significant *laterality* x *congruence* interaction called for a further simple effect analyses ($F(4,76) = 2.92, p = 0.03, \eta^2 = 0.13$). These analyses revealed that the effect of *congruence* was strongest in the right parietal cluster ($F(2,38) = 5.69, p = 0.01, \eta^2 = 0.23$) where a significant relative positivity compared to the *normal* condition was observed in both the *incongruous-resolved* condition ($F(1,19) = 4.77, p = 0.04, \eta^2 = 0.20$) and the *incongruous-unresolved* condition ($F(1,19) = 9.28, p = 0.01, \eta^2 = 0.33$). The effect of *congruence* was also significant in the midline parietal cluster ($F(2,38) = 5.03, p = 0.01, \eta^2 = 0.21$) where a significant positivity was seen only in the *incongruous-unresolved* condition ($F(1,19) = 8.91, p = 0.01, \eta^2 = 0.32$).

8.3.1.2 Discussion

For the *harmony* stimuli, in musicians only, the incongruity caused a significant positivity compared to the baseline (the *normal* stimulus) in both the *incongruous-resolved* and *incongruous-unresolved* stimuli in the right parietal region, but only between the *normal* and *incongruous-unresolved* in the midline parietal region. Since the *incongruous-resolved* and the *incongruous-unresolved* stimuli are identical up to the point of resolution, the same patterns were expected in both these conditions. However, because of the continuous nature of these auditory stimuli, the resolution in some stimuli happened before the end of the time-window in which the effect of the incongruity was analysed, making it possible for the continuation of the incongruity to affect the amplitude of the ERP effect observed in the incongruous stimuli.

The data shown in this section suggest that when the incongruity resolved, the amplitude of the ERP effect was smaller than when the incongruity did not resolve, and show similar patterns to the findings of Patel et al. (1998).

8.3.2 Rhythm

In this analysis, only 17 musicians achieved a sufficient number of trials in each condition cell to be included in the analysis. The waves for rhythm showed a slight positivity in the central and parietal clusters starting around 600ms and ending after 850ms. The time-window which best capture this difference between *congruence* conditions was 650-800ms. Though not discernible in the waveforms of the frontal electrodes due to noise, a frontal negativity was also observed during that time-window on the topographical maps (see Figure 8.3). Once again, when comparing musicians and non musicians, the effect seemed stronger in the former than in the latter.

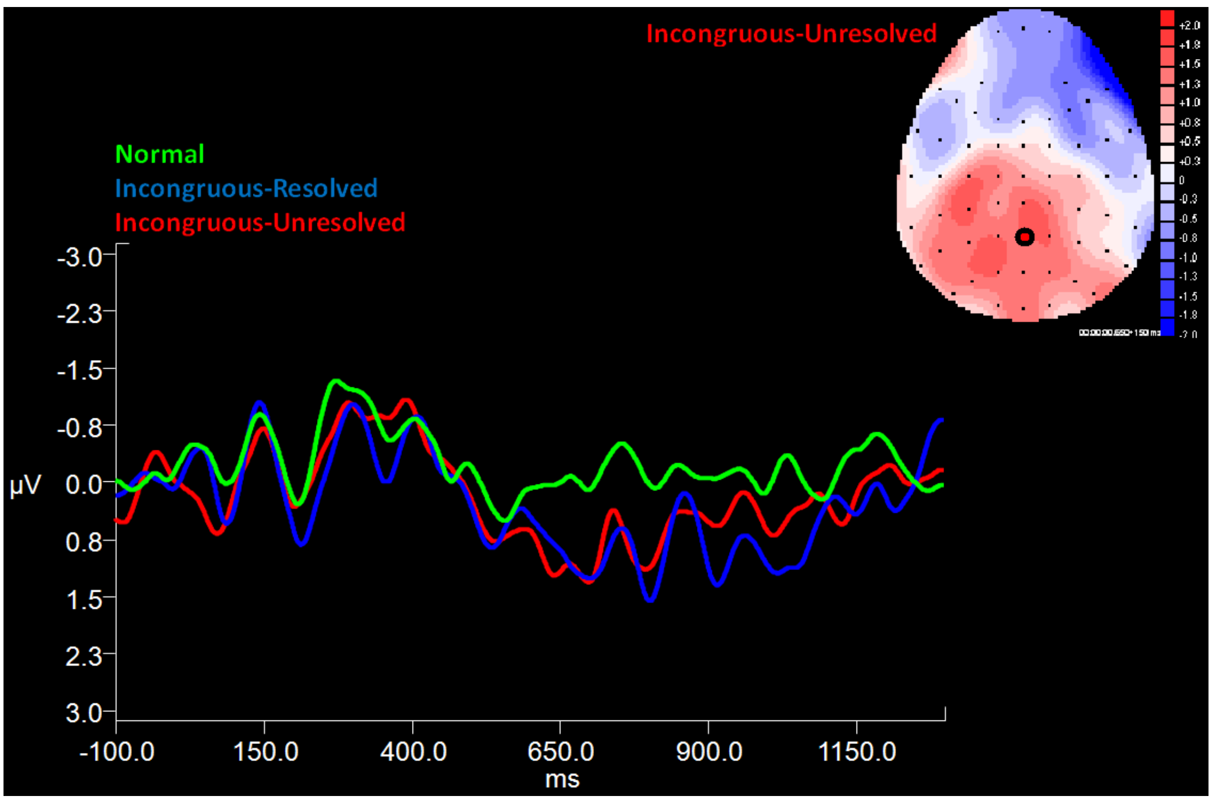

Figure 8.3: Waveforms for all three levels of *congruence* and topographical map for *incongruous-unresolved* minus *normal* in the 650 to 800ms time-window after the incongruity in *rhythm* across all participants

8.3.2.1 Results

The general ANOVA revealed no significant main effect for *congruence*, but a significant *congruence* x *laterality* effect ($F(4, 140) = 3.82, p = 0.01, \eta^2 = 0.10$), a significant *congruence* x *anterior-posterior* effect ($F(4, 140 = 4.21, p = 0.01, \eta^2 = 0.10$), a significant three-way interaction of *congruence* x *anterior-posterior* x *laterality* ($F8, 280 = 2.20, p = 0.04, \eta^2 = 0.06$) and a near significant three way interaction of *congruence* x anterior-posterior x musicianship ($F(4, 140) = 2.47, p = 0.08, \eta^2 = 0.07$). Follow-up simple analyses revealed no significant effect of *congruence* in the areas showing a slight positivity, but a significant effect of *congruence* in the frontal right cluster ($F(2, 72) = 3.33, p = 0.04, \eta^2 = 0.09$) showing a significant negativity in the *incongruous-unresolved* condition compared to the *normal* condition ($F(1, 36) = 5.52, p = 0.02, \eta^2 = 0.$). The *incongruous-resolved* condition also showed a negativity compared to the *normal* condition in this cluster, but the effect was not significant ($F(1, 36) = 2.66, p = 0.11, \eta^2 = 0.07$).

8.3.2.2 Discussion

It is thought that the noise observed in the frontal electrodes could have been due to participants tensing or frowning while listening to the rhythm stimuli. These suggestions were not supported by differences in the tension ratings between *rhythm* and the other levels of *hrss*. Participants were also observed tapping along to the rhythm stimuli. They were told to stop tapping as soon as this behaviour was noticed, and any trials in which participants had been tapping were rejected from the analysis due to the signal noise caused by this activity, but it could be that this tendency underlies some of the noise observed in this level of *hrss*. However, although these results must be interpreted with caution, these data do show a significant negativity in the frontal regions of the scalp following the onset of a rhythmic incongruity in the *incongruous-unresolved* condition compared to the *normal* condition in a region similar that reported by Münte, Nager, Beiss, Shroeder, and Altenmüller (2003). The noise observed in *rhythm* could also be due to between-participant variation in which meter participants adopted while listening to the stimuli, making it difficult to conclude as to the interpretation of the observed differences in ERP effects between conditions.

8.3.3 Semantics

When considering all participants together, the waveforms once more showed a strong overlap until approximately 300ms after incongruity onset at which point the stimuli in which the target was incongruous showed a relative long-lasting negativity, best captured by the time-window 350-750 ms after incongruity onset. Clear differences were apparent from looking at the waveforms and topographical maps separately for musicians and non musicians (see Figures 8.4 and 8.5). In the non-musician group, a right lateralised negativity was seen in the incongruous conditions compared to the *normal* condition, but this effect was not seen in musicians.

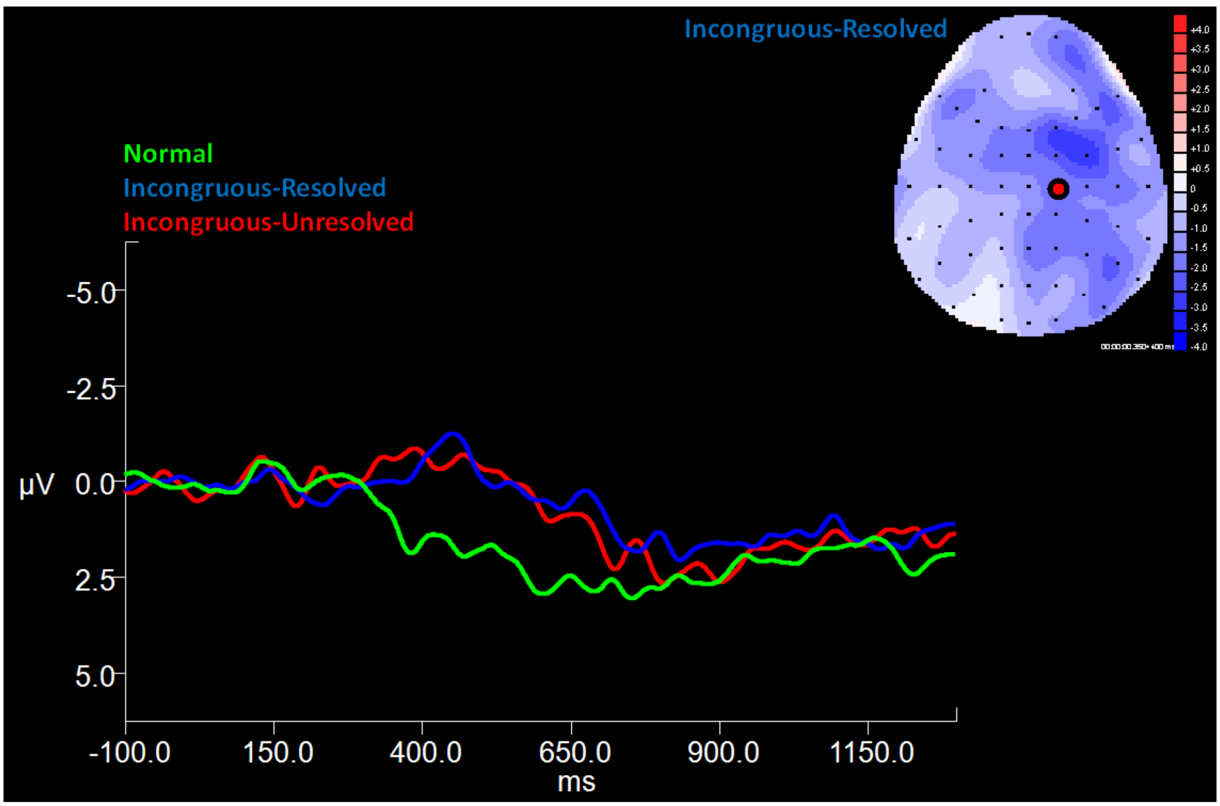

Figure 8.4: Waveforms for all three levels of *congruence* and topographical map for *incongruous-resolved* minus *normal* in the 350 to 750ms time-window after the incongruity in *semantics* in *non musicians*

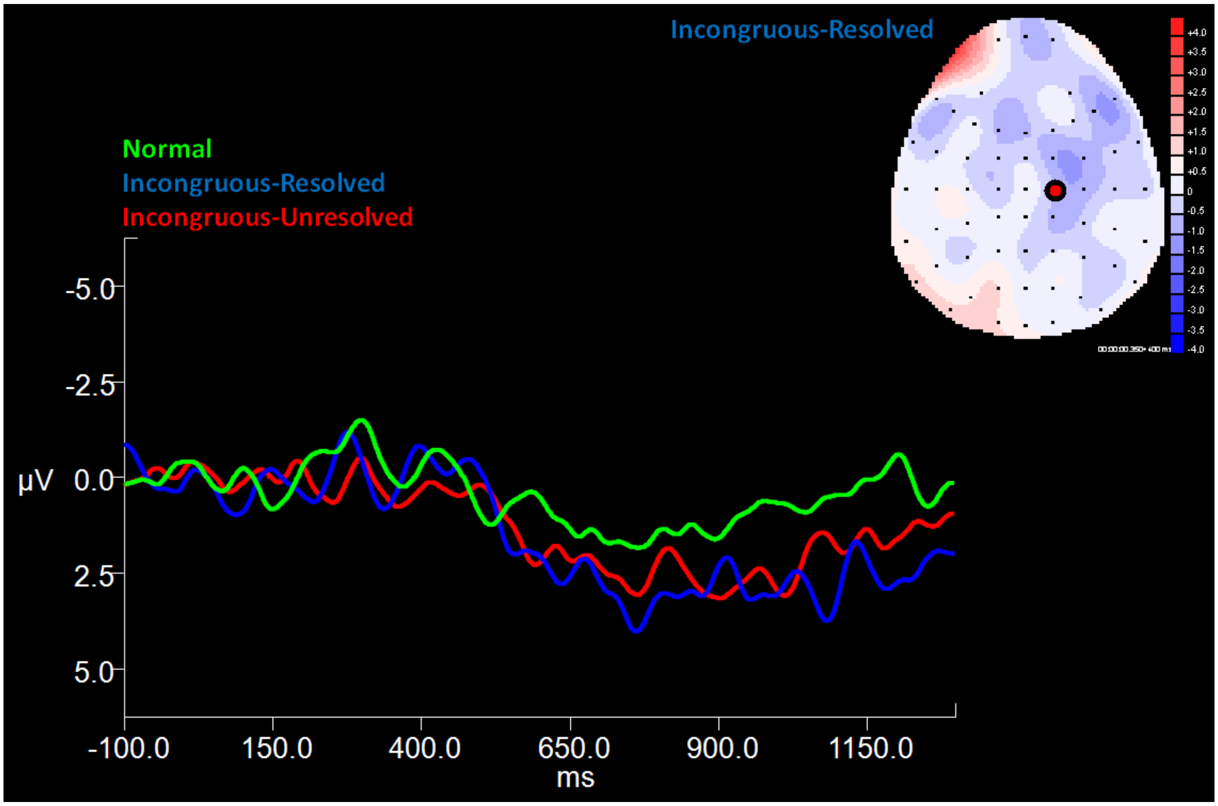

Figure 8.5: Waveforms for all three levels of *congruence* and topographical map for *incongruous-resolved* minus *normal* in the 350 to 750ms time-window after the incongruity in *semantics* in *musicians*

8.3.3.1 Results

In the general ANOVA for *semantics*, the only effect to approach significance was a four way interaction of *congruence* x *anterior-posterior* x *laterality* x *musicianship* ($F(8, 280) = 1.74, p = 0.08, \eta^2 = 0.05$). Considering the difference in patterns between musicians and non musicians, the simple effects analyses were carried out separately in each group. Unsurprisingly from the previous description of the data, no significant effects of *congruence* and no significant interaction effects of *congruence* with localisation factors were seen in musicians in this time-window. Thus the incongruity had no significant effect in musicians in the time-window typically associated with semantic incongruity-related ERP effects (all effects, $F < 1.50, p > 0.2, \eta^2 < 0.09$).

In the non-musician group, simple effects analyses revealed a significant main effect for *congruence* both across the midlines clusters ($F(2, 38) = 3.57, p = 0.05, \eta^2 = 0.16$)

and the right clusters ($F(2,38) = 5.968, p = 0.006, \eta^2 = 0.24$), with a significant negativity in response to the incongruity found in both the *incongruous-resolved* and *incongruous-unresolved* condition in comparison to the *normal* condition (respectively, midline: $F(1,19) = 4.98, p = 0.04, \eta^2 = 0.21$, $F(1,19) = 3.63, p = 0.07, \eta^2 = 0.16$; right: $F(1,19) = 9.31, p = 0.01, \eta^2 = 0.33$, $F(1,19) = 5.82, p = 0.03, \eta^2 = 0.24$).

To make sure that the lack of any effect in musicians was not due to a lack of power, 17 participants were selected randomly from the non-musician group to see whether a significant effect was still achieved with the same number of participants as in the musician group. A significant main effect for *congruence* was seen in the midline regions ($F(2,32) = 3.38, p = 0.05, \eta^2 = 0.17$) and in the right clusters ($F(2,32) = 4.99, p = 0.01, \eta^2 = 0.24$). The planned contrasts for these effects were as follows, presented respectively for *incongruous-resolved* and *incongruous-unresolved* compared to *normal*: midline $F(1,16) = 4.02, p = 0.06, \eta^2 = 0.02$, $F(1,16) = 4.80, p = 0.04, \eta^2 = 0.23$; right: $F(1,16) = 7.00, p = 0.02, \eta^2 = 0.30$, $F(1,16) = 6.72, p = 0.02, \eta^2 = 0.30$.

8.3.3.2 Discussion

These analyses revealed a significant negativity in response to semantically incongruous elements in non musicians only, in the form of a slightly right lateralised effect distributed from parietal to frontal regions similar to the effects reported in previous studies to do with the processing of metaphorical words and nonsensical sentences. The fact that the effect in the *incongruous-resolved* condition narrowly failed to reach significance in the midline regions in this analysis is in line with previous findings of a more right lateralisation of effects when participants process metaphoric incongruities (Coulson & Van Petten, 2007). In both region clusters, the difference between *incongruous-resolved* stimuli and *normal* was stronger than the difference between *incongruous-unresolved* and *normal*, in line with previous research in this area (Tartter et al., 2002; Bonnaud et al., 2002; Coulson & Van Petten, 2002; Kazmerski et al., 2003). The longer latency of the metaphorical incongruity compared to the nonsensical incongruity also replicates previous findings (Coulson & Van Petten, 2007). The lack of any significant effect of *congruence* on ERP patterns in musicians during this time-window was surprising. Several tentative explanations for this unexpected finding are offered in the Discussion section of this chapter and in the General Discussion of this thesis.

8.3.4 Syntax

In this analysis, two participants were excluded from the musicians group for failing to provide enough good trials in each condition cell. Observation of the waveforms revealed two areas of departure in stimuli containing incongruities. The data presented in Figure 8.6 show a departure between 300-500 ms, with a relative negativity for the incongruous stimuli compared the the *normal* stimuli over the midline and left clusters. A similar pattern has been found in syntactic violations where the word constituting the violation has a high semantic content (Luck, 2006; Phillips, Kazanina, & Abada, 2005). The second departure from the *normal* condition was a relative positivity between 800 and 1000ms. This could be the late positivity typically associated with syntactic incongruities, the effect of the word-class violation not taking full effect until its semantic content has been processed.

8.3.4.1 Results

The overall ANOVA on the 300-500 time window revealed a significant main effect for *congruence* ($F(2, 72) = 4.37, p = 0.02, \eta^2 = 0.11$), a marginally significant *congruence* x *laterality effect* ($F(4, 144) = 2.40, p = 0.07, \eta^2 = 0.06$) and no significant interaction effects with musicianship (all $F < 1$). Since the interactions with musicianship were far from significant, and since no differences were observed in the waveforms and topographical maps between the two subject groups, both subject groups were analysed together in follow-up analyses.

As suggested by Figure 8.6, simple effects analyses revealed a significant effect for *congruence* throughout the left clusters ($F(2, 74) = 4.47, p = 0.02, \eta^2 = 0.11$) with

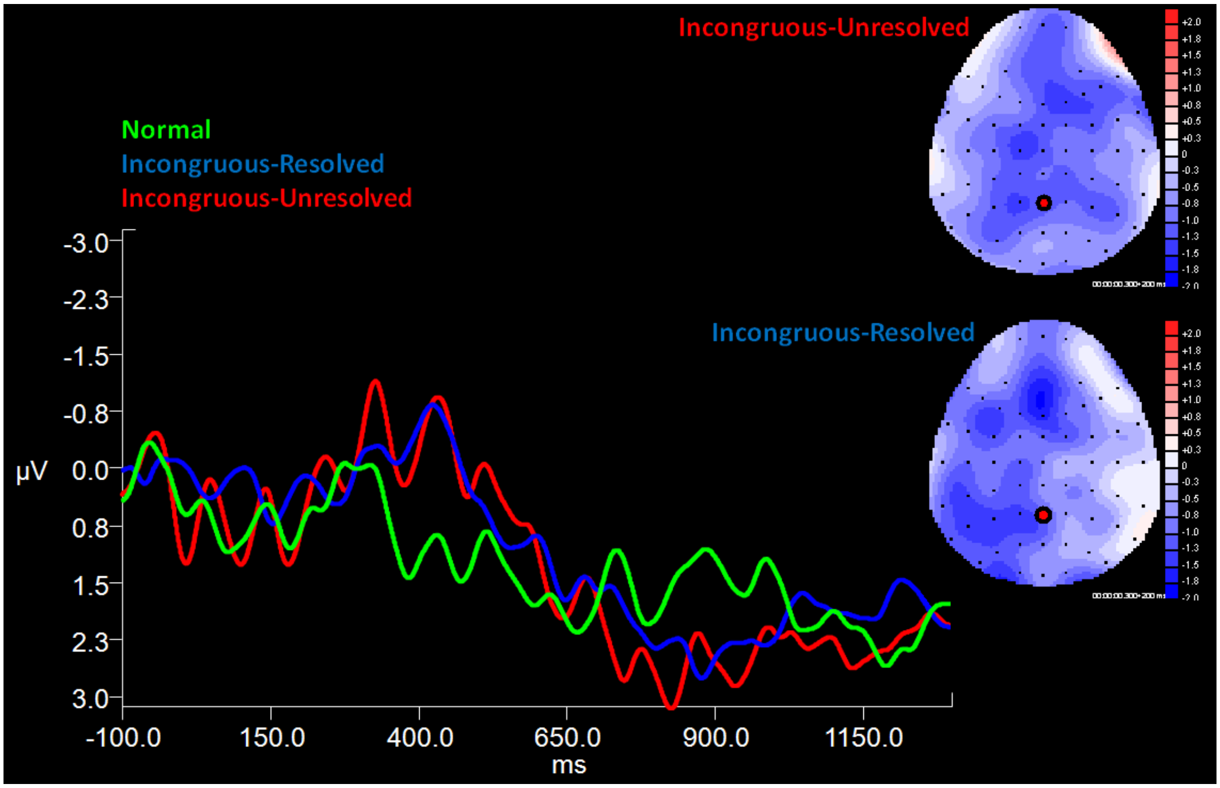

Figure 8.6: Waveforms for all three levels of *congruence* and topographical maps for the 300 to 500ms time-window after the incongruity for *syntax* across all participants

a significant negativity in both *incongruous-resolved* and *incongruous-unresolved* in comparison to *normal* (respectively $F(1,37) = 7.64, p = 0.01, \eta^2 = 0.17$ and $(F(1,37) = 3.98, p = 0.05, \eta^2 = 0.10)$. The pattern was similar for the midline clusters, with a significant main effect for *congruence* $(F(2,74) = 4.91, p = 0.01, \eta^2 = 0.12)$ with again a significant departure in both *incongruous-resolved* and *incongruous-unresolved* (respectively $F(1,37) = 7.06, p = 0.01, \eta^2 = 0.16$ and $F(1,37) = 5.68, p = 0.02, \eta^2 = 0.13$). The main effect for *congruence* narrowly failed to reach significance in the right clusters $(F(2,74) = 2.92.p = 0.06, \eta^2 = 0.07)$, confirming that the effect was slightly left lateralised.

When the effect of *congruence* was observed within each level of *laterality*, a significant effect for *congruence* was seen in the left regions $(F(2,74) = 4.47, p = 0.02, \eta^2 = 0.11)$ with no significant interaction with anterior-posterior, and a significant departure from the *normal* condition in both the *incongruous-resolved* $(F(1,37) = 7.64, p = 0.01, \eta^2 = 0.17)$ and the *incongruous-unresolved* condition $(F(1,37) = 3.98, p = 0.05, \eta^2 =$

0.10). Midline regions also showed a significant main effect for *congruence* ($F(2,74) = 4.91, p = 0.01, \eta^2 = 0.12$) with again a significant departure from *normal* in both *incongruous-resolved* and *incongruous-unresolved* (respectively $F(1,37) = 7.06, p = 0.01, \eta^2 = 0.16$ and $F(1,37) = 5.68, p = 0.02, \eta^2 = 0.13$), stronger in the *incongrous-unresolved* condition than in the left regions. The effect was only marginally significant in the right regions ($F(2,74) = 2.92.p = 0.06, \eta^2 = 0.07$) with a significant departure only in the *incongruous-unresolved* condition ($F(1,37) = 5.29, p = 0.03, \eta^2 = 0.13$). This suggests a more left-lateralised effect in resolved incongruities and a more right-lateralised effect in unresolved incongruities.

In the 850-1000 time-window, no significant main effect of *congruence* was seen ($F < 1$) but the *congruence* x *anterior-posterior* interaction and the *congruence* x *anterior-posterior* x *laterality* interaction were both significant (respectively $F(4,144) = 3.06, p = 0.05, \eta^2 = 0.0.08$ and $F(8,288) = 2.54, p = 0.02, \eta^2 = 0.07$). Simple effects analyses did not find a significant main effect for *congruence* in any of the clusters or combinations of clusters suggested by the interactions, meaning that the effect of the incongruity within this second time window was not significant.

8.3.4.2 Discussion

The data for the syntactic incongruities showed a significant relative negativity in the 300-500 time-window across the midline and the right regions in response to the incongruity in both *incongruous-resolved* and *incongruous-unresolved* conditions compared to the *normal* condition.

Although a positivity was expected in this condition following previous findings using a functional shift (Thierry et al., 2008), a significant negativity has been seen in response to syntactic incongruities when the semantic content of the incongruity is high (e.g. a noun compared to a preposition). Phillips et al. (2005) found a sustained negativity in the 300 to 500ms time-window after the onset of a verb constituting a grammatical difficulty in *wh-* dependency sentences such as *The lieutenant knew* *which accomplice the detective* ***hoped*** *that the shrewd witness would recognise in the lineup*. The authors suggested that this negativity was due to the verb preventing the immediate completion of the *wh-* dependency. Lee and Federmeier (2006) demonstrated a negativity in a 250 to 900ms time-window in response to word-class violations only when the violation led to semantic ambiguity and required semantic resolution. Since the functional shifts used throughout this thesis were not ones which had passed into common usage (e.g. to refrigerate the food, to butter the bread, to hoover the floor), it could be that their novel usage required further semantic processing to make sense of the sentence. This, in turn, could underlie the evocative nature of the syntactic shift (Clark & Clark, 1979; Thierry et al., 2008), though no increased aesthetic experience was recorded in the *syntax* stimuli in this experiment. The slight relative positivity observed in the later time-window failed to reach significance. The discussion of the analysis of the point of resolution in syntax addresses this issue.

8.3.5 Summary of ERP effects at the point of incongruity

At the point of incongruity, the data showed a significant positivity around the time-window of the P600 (500 to 700ms) in musicians in *harmony*, similar to the pattern observed in Patel et al.'s (1998) seminal study. In *rhythm*, the posterior positivity was not significant, but a significant negativity was found in the frontal areas in the 650 to 800ms time-window in a similar region to the one showing an effect of rhythmic incongruities in previous research (Münte et al., 2003). In semantics, non musicians showed a pattern similar to that found in previous studies investigating metaphor processing (Coulson & Van Petten, 2007), namely a significant right-lateralised sustained negativity in the 350 to 750 time-window. Surprisingly, musicians showed no significant ERP effect at the point of incongruity, suggesting an effect of musical training on language processing. Possible explanations for this observation are discussed in the General Discussion of the thesis (Chapter 9). In *syntax*, the positivity reported by Thierry et al. (2008) in response to a functional shift was not found, however a significant sustained negativity resembling the effect found by Phillips et al. (2005) and Lee and Federmeier (2006) suggested that the semantic processing of the word which constituted a word class violation was predominant over the syntactic integration required by this incongruity.

8.4 ERP effects at the point of resolution, or of "confirmation of nonsense"

The stimuli in this project were developed with the purpose of both being able to investigate the effect of computationally equivalent incongruities in all four components of music and language, but also to look at the effect of resolution — an aspect which is central to theories of aesthetics but which has not been studied empirically to date.

In the *normal* stimuli, since the context matches the element of interest, the point of resolution is merely a normal continuation in the sentence. In the *incongruous-resolved* stimuli, the resolution is the point at which the stimulus returns to the original context (e.g. the first note indicating a return from the original key). In the *incongruous-unresolved* condition, this point could be dubbed the point of "confirmation of nonsense", the point at which it is no longer possible to integrate the previously ambiguous element into the sentence or the piece. To investigate the ERP patterns associated with the point of resolution (or in the case of unresolved stimuli, the point at which a resolution becomes impossible), the same approach was taken, analysing the data time-locked to the point at which the new key or beat was confirmed in music, and the word after the syntactic incongruity in language.

8.4.1 Harmony - P600 in musicians, N4/500 in non musicians

The harmony data showed two different effects in musicians and in non musicians. The former had shown a late positivity, in the time-zone of the P600, in response to the incongruity. In response to the lack of resolution, another late positivity was found.

In non musicians, who showed no significant ERP effect at the point of incongruity, a negativity was seen upon the lack of resolution.

8.4.1.1 Musicians

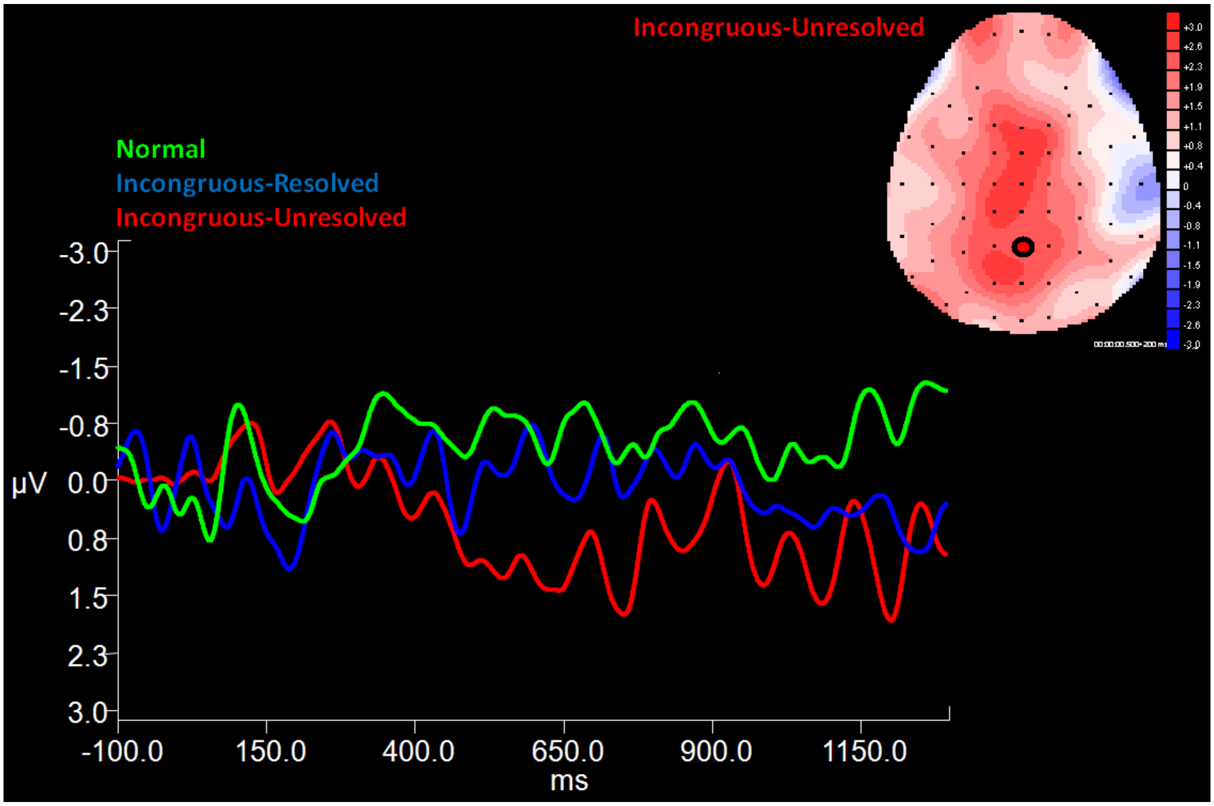

Figure 8.7: Waveforms for all three levels of *congruence* and topographical map for *incongruous-unresolved* minus *normal* in the 500 to 700ms time-window after the point of resolution in *harmony* in *musicians*

Figure 8.7 displays the topographical map of the difference between the *normal* and the *incongruous-resolved* condition. The ANOVA carried out to assess the significance of this effect showed a significant *congruence* x *laterality* interaction effect ($F(4, 76) = 4.13, p = 0.004, \eta^2 = 0.18$). Follow-up simple effects analyses revealed that the effect of *congruence* was significant in the middle regions ($F(2, 38) = 4.54, p = 0.02, \eta^2 = 0.19$). As suggested by the waveforms shown in Figure 8.7, the planned contrasts in this region showed a significant positivity in the *incongruous-unresolved* condition relative to the *normal* condition ($F(1, 19) = 6.43, p = 0.02, \eta^2 = 0.25$). No significant difference was seen between the *incongruous-resolved* and the *normal* conditions ($F(1, 19) = 1.33, p =$

$0.26, \eta^2 = 0.07$). These data indicate that the lack of resolution of the incongruity (or the confirmation of a new key dictated by the incongruity) was perceived as a new incongruous element while the return to the original key was not, possibly because musicians had already interpreted the first incongruity as a chord borrowed from a nearby key which did not necessitate a full key modulation.

8.4.1.2 Non musicians

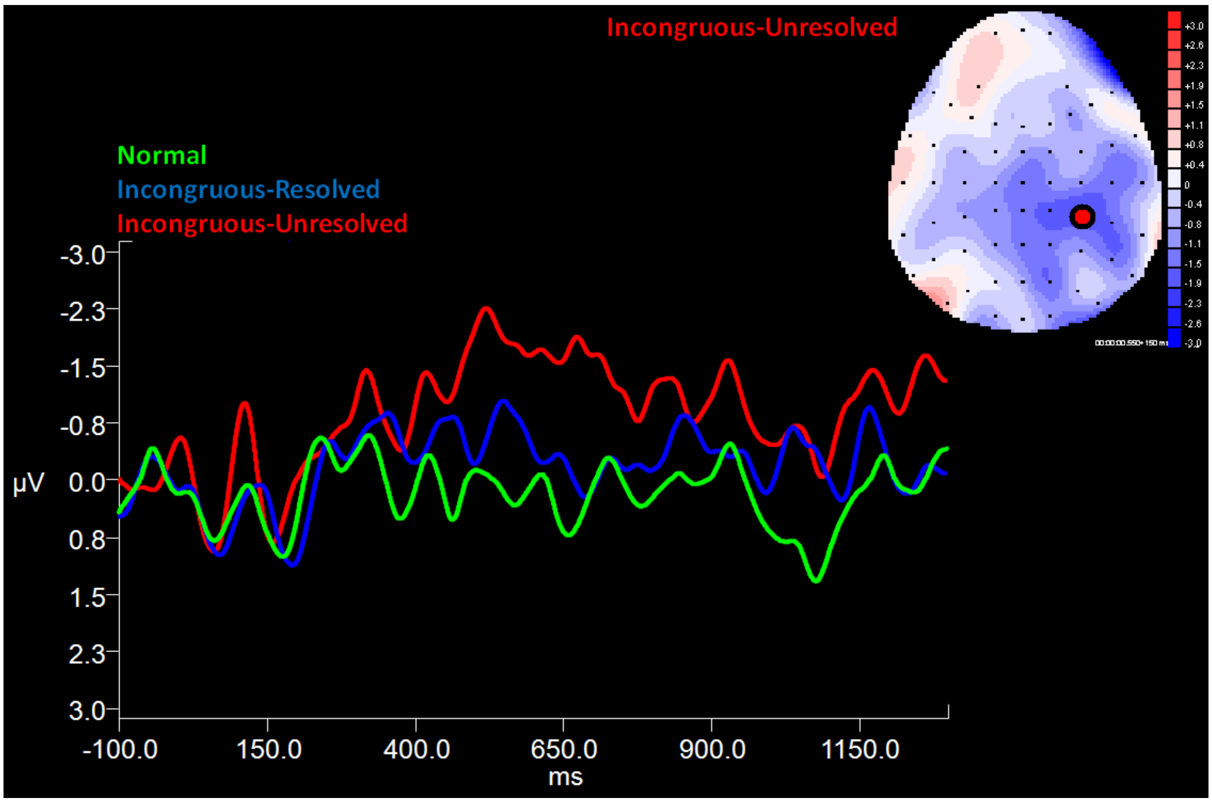

Figure 8.8: Waveforms for all three levels of *congruence* and topographical map for *incongruous-unresolved* minus *normal* in the 550 to 700ms time-window after the point of resolution in *harmony* in *non musicians*

In non musicians, who had not shown any significant ERP effect upon the original incongruity, a late negativity was found, in the region of the N400/500 reported in studies by Koelsch and his colleagues. Figure 8.8 shows the topographical map for the difference between *normal* and *incongruous-unresolved* from 550 to 700ms after the incongruity. Though the effect began around 400ms, the effect was sustained, and the maximum amplitude and most consistent topographical distribution was seen in this time-window.

Though effects in the general ANOVA failed to reach significance, the analysis of the effects in the right-parietal region in which the negativity was seen in Figure 8.8, showed a near-significant effect for *congruence* ($F(2, 38) = 2.79, p = 0.08, \eta^2 = 0.13$) and a significant relative negativity in the *incongruous-unresolved* condition compared to the *normal* condition ($F(1, 19) = 6.14, p = 0.02, \eta^2 = 0.24$). No significant difference was found between *incongruous-resolved* and *normal*, suggesting again that the return to the original key was not perceived as an incongruity. In the midline parietal cluster also, planned contrasts showed a significant relative negativity in the *incongruous-unresolved* condition compared to the *normal* condition ($F(1, 19) = 7.07, p = 0.02, \eta^2 = 0.24$) despite a non-significant effect of *congruence* in the ANOVA for that cluster ($F(2, 38) = 2.26, p = 0.12, \eta^2 = 0.11$). These data indicate that although non musicians showed no significant ERP effects in response to subtle incongruities, the lack of resolution of these incongruities led to an N4/500-type effect similar to the effect reported by Koelsch et al. (2001).

8.4.1.3 Discussion: confirmation of the role of closure in distinguishing between negativities and positivities in response to harmonic incongruities

These data help elucidate why the different research teams working in the field of music ERP research have found different responses to harmonic incongruities. In studies by Koelsch and his colleagues, the incongruity is typically placed at the end of a musical sequence and prohibits the correct resolution of the musical phrase. In studies by Patel and his colleagues, the incongruity is typically placed in the middle of the sequence in the place of a chord upon which the resolution of a phrase does not depend. These findings also lend support to the explanation reached through the systematic review carried out of the onset of this project, namely that the N400/500 in music is a reflection of a lack of resolution rather than the consequence of a statistically improbable chord per se.

8.4.2 Rhythm

In *rhythm*, the topographical map displayed in Figure 8.9 showed a late relative negativity in response to *incongruous-resolved* stimuli in the frontal areas of the scalp, with the clearest effect being from 750-900ms. No differences were observed between the waveforms for the *normal* and the *incongruous-unresolved* condition.

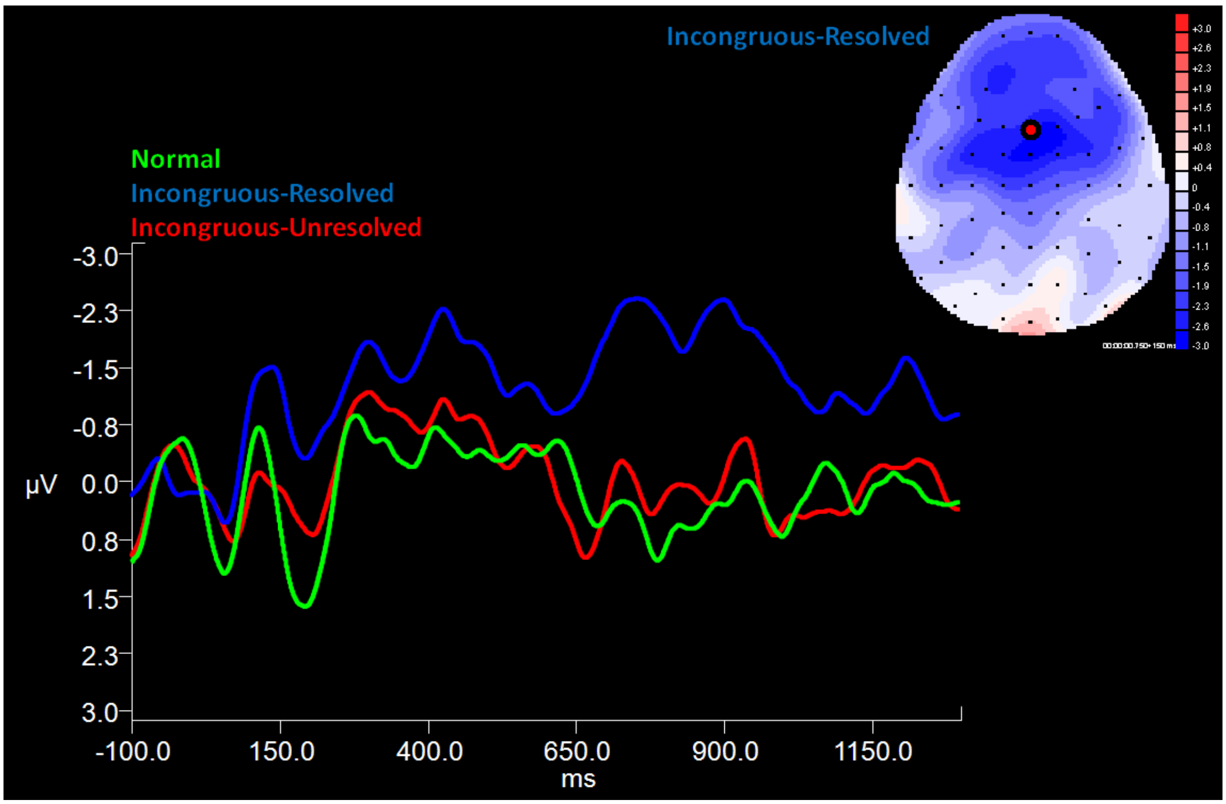

Figure 8.9: Waveforms for all three levels of *congruence* and topographical map for *incongruous-resolved* minus *normal* in the 550 to 700ms time-window after the point of resolution in *rhythm* across all participants

8.4.2.1 Results

The overall ANOVA showed a significant effect for *congruence* ($F(2, 74) = 5.21, p = 0.006, \eta^2 = 0.128$), and significant interactions of *congruence* with *laterality* ($F(4, 148) = 4.52, p = 0.002, \eta^2 = 0.11$) and *anterior-posterior* ($F(4, 148) = 5.66, p < 0.001, \eta^{=}0.13$), calling for further simple effects analyses. These revealed that the effect of *congruence* was strongest in the midline frontal cluster ($F(2,76) = 9.19, p < 0.001, \eta^2 = 0.20$) where a significant negativity was observed in the *incongruous-resolved* condition compared to the *normal* condition ($F(1,38) = 18.68, p < 0.001, \eta^2 = 0.33$). Contrary to the patterns observed in *harmony*, these data indicate that the return to the original beat elicited an incongruity-type response similar to and stronger than the effect elicited by the onset of the incongruous beat. No significant differences were seen between the *incongruous-unresolved* and the *normal* condition ($F(1,38) = 0.62, p = 0.44, \eta^2 = 0.02$), suggesting that the confirmation of the new beat did not lead to an incongruity-type response. These results must be interpreted with caution in the light of previous discussions concerning the maintaining of meter during rhythm perception, but they do seem to indicate a different time-course in rhythmic integration to that observed in the processing of *harmony* stimuli.

8.4.2.2 Discussion: Rhythmic resolution as a second incongruity

The noise discussed in relation to the *rhythm* ERP effects and the difficulty in establishing how participants perceived the rhythmic sequences in terms of the adoption of a stable meter make the interpretation of *rhythm* incongruity effects. Nonetheless, these data seem to confirm the speculations made upon observing the behavioural data relating to oddity ratings in rhythm: both the incongruous element and its resolution (the departure from and the return to the initial rhythmic pattern) were perceived as incongruities. This suggests that the time it takes to perceive a new pattern as the stable norm is much shorter in rhythm than in harmony. The stimuli I have created would allow the investigation of the effect of incongruity duration. The ERP effects seen upon resolution called for another experiment to be run in which enough iterations of each incongruity duration were used. This was not possible in the present experiment due to the amount of time it would have taken to run the experiment. The fact that both musicians and non musicians showed a significant incongruity-type ERP effect upon the resolution, and not upon a "confirmation of nonsense" suggests that this perception is independent from musical training.

8.4.3 Semantics

Pinpointing the exact moment at which a sentence begins to make sense again is very difficult. Subsequent studies using self-paced reading and eye-tracking need to be carried out before the semantic resolution data can be analysed in any meaningful way.

Further studies using self-paced reading or eye-tracking would need to be carried out to identify the word which forms the semantic resolution in each sentence before the data could be averaged across the resolution points in this level of *hrss*. However, since a trigger was placed at the onset of each word following the incongruity in *semantics*, once the point of resolution has been established in each sentence, it would be possible to use the data from this experiment to investigate ERP patterns associated with semantic resolution. Such studies will form the basis of research proposals for post-doctoral work, and would serve the double purpose of investigating differences between musicians and non musicians in the processing of semantic incongruities, as well as enabling further analysis of the ERP data gathered during this thesis.

8.4.4 Syntax

Although the same issue could be raised in the case of syntactic incongruities, the word after the incongruity was considered to provide a similar type of information in each stimulus, in view of the shared syntactic structure of the region of interest in these sentences. In the phrase *Mrs Matthews trousers her son*, the word *her* is the beginning of the object, confirming that the word *trousers* should be interpreted as a verb, not a noun. Thus, although this word does not provide syntactic resolution at the level of the sentence, it helps participants resolve the syntactic ambiguity concerning the word-class of the word of interest (e.g. *trousers*). The ERP effects on the word after the incongruity showed a relative positivity in the 250-750ms time-window in parietal regions of the scalp (see Figure 8.10), with the clearest pattern and most consistent difference between wave forms being found in the 300 to 500ms time window.

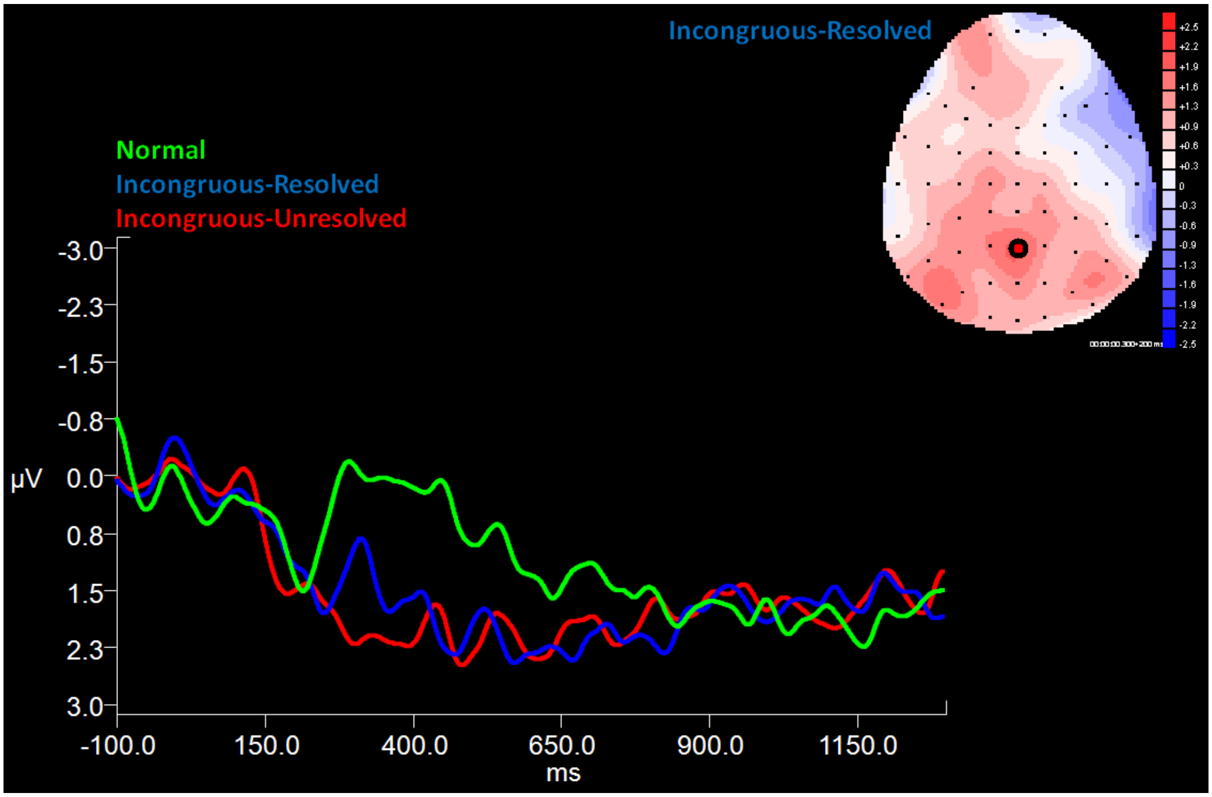

Figure 8.10: Waveforms for all three conditions and topographical map for *incongruous-resolved* minus *normal* in the 300 to 500ms time-window after the point of resolution in *syntax* across all participants

8.4.4.1 Results

The general ANOVA showed a significant *anterior-posterior* x *laterality* x *congruence* interaction effect ($F(8, 296) = 2.56, p = 0.03, \eta^2 = 0.07$) but no significant interaction with musicianship ($F(8, 296) = 0.19, p = 0.97, zeta^2 = 0.005$). Simple effect analyses revealed a significant effect for *congruence* in the midline parietal cluster ($F(2, 76) = 4.40, p = 0.02, \eta^2 = 0.10$), as suggested by Figure 8.10, with a significant relative positivity in the *incongruous-unresolved* condition compared to the *normal* condition ($F(1, 38) = 7.40, p = 0.01, \eta^2 = 0.16$). The positivity observed in the *incongruous-resolved* condition compared to the *normal* condition failed to reach significance ($F(1, 38) = 3.08, p = 0.09, \eta^2 = 0.08$).

8.4.4.2 Discussion

The distribution of this effect is similar to that of a P600 typically related to syntactic processing. Its odd timing could be due to two things. Firstly, it could be a late P600 in response to the incongruous word (the previous trigger). This is however unlikely, since the time distance between this “resolution” trigger and the onset of the previous word would vary across stimuli. A weaker effect rather than a stronger effect of *congruence* should therefore be seen calculated in relation to the “resolution” trigger than in relation to the “incongruity” trigger, because of a noisier signal (the incongruity itself) when creating an average waveform for each condition. A second and more plausible explanation is that this effect is an early P600 in response to the word after the incongruity (the “resolution” trigger), showing a shorter onset latency than a typical P600 effect because of the speed at which the language is delivered in continuous speech in auditory stimuli compared to the typically slow paced visual presentation used in the majority of studies reporting a P600 effect in response to syntactic violations (Hagoort & Brown, 2000).

Since this effect is only significant in the *incongruous-resolved* condition while the words up to this point are identical in this condition and in the *incongruous-resolved* condition indicates that the amplitude of the effect must be linked to the difficulty of integrating it into the sentence. Indeed the average duration of the incongruous word in the *syntax* stimuli was 440.63ms, and the average duration of the first word after the incongruity was 222.47ms. The trigger for the word after the incongruity therefore occurred on average 440.63ms after the trigger for the incongruous word, making a strong reaction from 300 to 500ms after the first word after the incongruity a significant effect from 700 to 900ms after the incongruous word. Furthermore, the second word after the incongruity lasted on average 315.56ms. In this case, the onset of the 3rd word after the incongruity would take place within 600ms of having heard the first word after the incongruity, making it very plausible that it could affect the integration difficulty of the target word and therefore the amplitude of the P600-type effect observed here.

8.4.5 Summary of ERP effects at the point of resolution

At the point of resolution in *harmony*, the data first of all showed a significant negativity in non musicians in response to a lack of resolution where a significant positivity was found in musicians. This enables the resolution of the discrepancy between studies showing these two types of effects, since the effect typically found in non musicians is associated with an interrupted cadence in which an incongruous chord, placed without regard for musical good form, prevents the closure of the harmonic progression. In *rhythm*, despite the discussed difficulty in establishing resolution, the presence of a second incongruity-type component at the point of resolution only in *incongruous-resolved* stimuli suggests that the return to the original beat pattern is perceived as a second incongruity. This finding helps elucidate why oddity rating trends were not as expected in Experiment 1 and prompts the question of whether the temporal nature of resolution varies between *harmony* and *rhythm*. Since this issue is captured by the data from both Experiment 1 and Experiment 3, it is discussed in greater detail in the General Discussion of this thesis. Finally, in *syntax*, a significant positivity was seen in response to the work following the incongruity in the area of the scalp typically seen in a P600 effect associated with syntactic processing. This effect suggests that though the incongruity itself led to more semantic-type effects than syntax-type effects, the necessity of integrating the functional shift noun as a verb does result in syntax-like effects. The fact that the late positivity observed in the analysis of ERP effects associated with the point of incongruity was not significant lends support to the notion that the word following the functional shift elicits the syntax-type effects. Indeed the realigning of the waveforms with the precise onset of the word following the incongruity should, and did, lead to a strengthening of such an effect.

8.5 Adding ERP data to the equation

The data presented thus far in the thesis show a significant increase in ratings of aesthetic pleasantness in the *incongruous-resolved* condition, in *harmony* in the musician group and in *semantics* in the non-musician group. The analysis performed on the behavioural data suggested that level of reported stimulation and tension felt after listening to the stimulus were both significant predictors of aesthetic ratings in *harmony* and *semantics*. The differences between musicians and non musicians were however not explained by the effect of oddity, as both groups showed a similar pattern in oddity ratings in both these levels of *hrss*.

8.5.1 Data preparation and analyses

In order to examine the ERP data and the scale data side by side, the amplitude of the ERP component needed to be transformed into a value which could be compared to the values from the behavioural data. ERP effects are only meaningful in terms of differences between conditions. Since in this experiment the baseline was the *normal* condition, difference scores were computed to create an "amplitude of the ERP effect score" in the *incongruous-resolved* condition and in the *incongruous-unresolved* condition by subtracting the mean amplitude in the *normal* condition from the mean amplitude in each *congruence* condition and each level of *hrss* in each participant. This resulted in either a positive or a negative score in the *incongruous-resolved* and *incongruous-unresolved* condition, depending on the nature of the ERP effect, and a score of 0 in the *normal* condition.

To make the scores comparable to the behavioural data, and to eliminate the statistical problem associated with having a value of 0 as the mean for all the stimulus means in the *normal* condition, these difference scores were transformed into z-scores within each participant. Where the ERP effect was a negativity (in rhythm, semantics, syntax at the point of incongruity and harmony for non musicians at the point of resolution), this z-score was multiplied by -1. This final step allowed the investigation of whether the same relationship could be observed across all levels of *hrss*. Thus for each level of *congruence* and each level of *hrss*, each participant had one ERP amplitude value represented by a z-score and one z-transformed behavioural value.

Though an unorthodox approach to ERP analysis, this procedure, discussed with experts in the field of music and language ERP research, seemed to be the only way of meaningfully analysing the behavioural and the ERP data side by side. The results discussed in this section support the need to investigate these two types of data together in this experiment, and show greater insight into the underpinnings of the role of incongruities in aesthetics across domains.

These ERP effect amplitude z-scores were entered into a correlation with the data from the scales analysed in Chapter 7. The frequency distributions for the z-scores of the ERP components showed an inverted U-curve pattern, with higher frequencies at the extremities of the spectrum than in the middle of the spectrum. In view of this violation of the normality parameter, a Spearman's rank order correlation was chosen for these analyses. Considering the common assumption that the amplitude of an ERP component is a reflection of the degree of oddity or processing difficulty associated with the stimulus, these correlations were one-tailed.

8.5.2 Correlations across all levels of *hrss* and within each level of *hrss*

When all levels of *hrss* were pooled together, the strongest correlation observed between the amplitude of the component at the onset of the incongruity and the behavioural data was with oddity ratings ($r = .13, N = 474, p = 0.002$). The only other significant correlations with this ERP amplitude were the rating of how tense participants felt while listening to the stimuli ($r = .10, N = 474, p = 0.01$) and how confused as a result of hearing the stimuli ($r = 0.09, N = 474, p = 0.03$). At the point of resolution, for which no data from *semantics* was included (see Section 8.4.3), the amplitude of the ERP effect was also significantly correlated with *oddity* ratings ($r = .18, N = 348, p < 0.001$), confusion ratings ($r = .10, N = 348, p = 0.03$) and *tension during listening* ratings ($r = .09, N = 351, p = 0.05$). These data indicate that greater *oddity*, *tension during listening*, and *confusion* ratings were associated with a greater ERP effect amplitude, supporting the assumptions made in previous studies in this field of research (Jentschke et al., 2005; Kaan et al., 2000; Patel, 2003; Steinbeis et al., 2006).

8.5.2.1 Harmony

Within *harmony*, the rating scales in which the data correlated significantly with the amplitude of ERP effects at the point of incongruity were *oddity* ($r = 0.20, N = 120, p = 0.02$), *stimulation* ($r = 0.16, p = 0.05$) and *tension during listening* ($r = .15, N = 120, p = 0.05$). All three correlations were positive, as predicted, meaning that when participants showed a significant ERP effect at the point of incongruity, they reported a higher rating of oddity, tension and stimulation. At the point of resolution, the ratings which showed a significant correlation with the amplitude of the ERP component were *oddity* ($r = 0.20, N = 117, p = 0.02$) and *tension during listening* ($r = 0.20, N = 120, p = 0.02$), with the correlation with *confusion* failing to reach significance ($r = 0.14, N = 117, p = 0.06$).

8.5.2.2 Semantics

Within *semantics*, the same patterns were observed with the significant correlations between the amplitude of the ERP effect and ratings of *tension during listening* ($r = .18, N = 117, p = 0.02$), *oddity* ($r = 0.17, N = 120, p = 0.03$) and *stimulation* ($r = .15, N = 117, p = 0.05$). Once again, all three correlations were positive, meaning that when participants showed a significant ERP effect at the point of incongruity, they reported a higher rating of oddity, tension and stimulation.

8.5.2.3 Rhythm

Within *rhythm*, the amplitude of the ERP effect at the point of incongruity was not significantly correlated to *oddity* ratings ($r = 0.03, N = 120, p = 0.39$). Instead, significant negative correlations were found with ratings of *interest* ($r = -.16, N = 120, p = 0.04$) and *stimulation* ($r = -.18, N = 117, p = 0.03$). Since these correlations are in the opposite direction to the predictions, they cannot be considered significant in this analysis. A two-way Spearmans' correlation would have found them non-significant. These patterns could however be due to the difficulty in establishing the significance of the ERP effect at the point of incongruity due to the noise seen in the data. At the point of resolution however, where the ERP effect was more clearly defined, the amplitude of the effect was significantly correlated with *oddity* ratings alone ($r = .24, N = 117, p = 0.005$). Within *syntax*, no significant correlations were observed with the ERP amplitude, either at the point of incongruity or the point of resolution.

8.5.3 A cross-domain incongruity-based model of aesthetics?

In the groups showing a significant increase in *aesthetic pleasantness* in the *incongruous-resolved* condition, no significant difference between the *incongruous-resolved* condition and the *normal* condition was seen in either the ratings of *stimulation*, the ratings of *tension during listening* or the patterns in the ERP data.

In *harmony*, non musicians showed a significant increase in *aesthetic pleasantness* ratings and no significant ERP effect at the point of incongruity. In this group no significant effect of *condition* was found in either *stimulation* or *tension during listening* ratings, while the musician group did show a significant effect of *congruence* in both those ratings, however the *tension during listening* ratings showed a stronger increase from *normal* to *incongruous-resolved* than from *incongruous-resolved* to *incongruous-unresolved*. In *semantics*, musicians showed a significant increase in *aesthetic pleasantness* rating and no significant ERP effect at the point of incongruity. In this group, no significant effect of *congruence* was found in *stimulation* ratings, where a significant effect was found in non musicians, and a significant increase was seen in the ratings of *tension during listening* from *normal* to *incongruous-resolved* but not from *incongruous-resolved* to *incongruous-unresolved*.

Although at first glance these data seem to suggest a model in which the effect of the incongruity is modulated by the presence or absence of an ERP effect which in turns optimises levels of tension and stimulation to achieve a heightened aesthetic experience, a closer observation of the data shows a more complex picture. Firstly it seems that to achieve a heightened aesthetic experience, an incongruity embedded in a stimulus must elicit an optimal level of stimulation which is slightly increased from the *normal* condition but not significantly so. Since the amplitude of the ERP effect was significantly positively correlated with the stimulation ratings, the absence of an ERP effect in the musicians in semantics and in the non musicians in harmony would achieve this effect. However, the fact that stimulation was a significant predictor of aesthetic pleasantness with a positive coefficient suggests that a further increase in stimulation would serve to increase the aesthetic pleasantness of the stimulus. Secondly, the fact that the *tension during listening* ratings showed a significant positive correlation with the amplitude of the ERP component was also confusing. The participants who showed a significant increase in aesthetic pleasantness ratings also showed a significant increase in tension during listening ratings, but no significant ERP effect.

Thus, though the way in which the participants' mind perceives and processes the incongruous element seems key to understanding the role of incongruities in aesthetic-affective responses to both music and language, no clear mechanisms could be deduced from the patterns explored here.

8.6 Discussion

The ERP effects reported in this chapter were similar to those reported in the literature in the fields of music and language incongruity processing. An important finding was the eliciting of a P600 at the point of incongruity in musicians, and an N400/500 at the point of lack of resolution in non musicians. This finding supports the role of closure in distinguishing between these two patterns, ventured in the discussion of ERP effects undertaken in the systematic review.

The cross-over noticed in the data reported in Chapter 7 suggested that musicians and non musicians responded differently to incongruities in *harmony* and *semantics*, despite similar trends in oddity ratings between musicianship groups. Adding ERP data to the model showed yet another cross-over in the data, with only the group in which no significant incongruity-locked ERP effect was elicited showing a significant increase in aesthetic ratings for *incongruous-resolved* stimuli.

The analysis of correlations between the amplitude of the ERP component and the data on the behavioural scales first of all supported the interpretation of ERP components as indicators of oddity detection and processing difficulty, and secondly demonstrated significant correlations between stimulation ratings and ERP components and tension during listening ratings and ERP components. Since stimulation and tension were shown to be predictors of aesthetic ratings, an attempt was made at a model which could capture the data trends explored thus far.

The clear pattern emerging from the data is that only a stimulus in which a resolved incongruity elicits an increased oddity rating which is not accompanied by an incongruity-locked ERP component results in a higher rating of aesthetic pleasantness than a stimulus which is not rated as odd. The question remains as to what causes different patterns to be observed in musicians and non musicians. This question is addressed in the General Discussion of this thesis (Chapter 9).

Chapter 9

General discussion: where weird is wonderful

9.1 Introduction

The research presented in this thesis was designed to test the notion that a shared model could account for the role of incongruities in eliciting aesthetic-affective responses to music and language. Alongside this main research endeavour, additional findings of relevance to the study of music and language cognition emerged, with implications for the interpretation of previously published studies, and suggestions for avenues of future research.

9.2 Main findings and implications from this research

The theoreties from music and language aesthetics discussed in the General Introduction (Chapter 1) and the outcomes of the systematic research review (Chapter 2) provided a foundation for the construction of a series of research questions answered below. As predicted in the introductory chapters, the research carried out in this thesis has implications beyond the fields of music and language incongruities and aesthetics.

9.2.1 The psychological literature supports the plausibility of a role for incongruities in aesthetic-affective responses to both music and language

The systematic literature review carried out at the outset of this research was based on the observations of striking similarities between theories of music aesthetics and language aesthetics reported in the General Introduction (Chapter 1). This review led to the conclusion that the similarities seen between music and language, in terms of the stimuli themselves, the way in which they are processed, the brain areas involved in this processing and the effect of incongruous elements, warranted their parallel investigation. These findings, combined with the call for a more systematic investigation of the role of incongruities in eliciting affective-aesthetic responses in both domains (Thierry et al., 2008; Steinbeis et al., 2006; Juslin & Västfjäll, 2008) and the observation of methodological and theoretical discrepancies within these fields of research, enabled the construction of further research questions based upon previous observations in the empirical literature in music and language psychology.

9.2.2 A stimulus set was developed, normed and validated for the purpose of studying incongruity processing with carefully controlled equivalent levels of congruence across music and language

In order to establish whether any common patterns could account for heightened aesthetic experiences in both domains, it was important to create stimuli which violated statistical probabilities in equivalent ways, and in which an identical element could be compared across conditions. This was achieved by starting from incongruities which had been used by composers and authors for aesthetic effect, in a manner akin to

Steinbeis et al. (2006), and modifying the context of this aesthetically purposed incongruity to maintain an identical element of interest across *congruence* conditions. The manipulation of the degree of incongruity as well as the degree of resolution enabled the investigation, detailed below, of whether the differences observed between the stimuli used by different research teams could account for the differences observed in ERP responses to music and language incongruities (see Tables 2.1 and 2.2).

The resulting stimulus set led to a significant increase in oddity ratings from *normal* to *incongruous-resolved* and from *incongruous-resolved* to *incongruous-unresolved* in all four components of music and language. The norming and validation of these stimuli, presented in Chapers 5 and 7, supported the plan to make these stimuli available to other researchers within music psychology, language psychology, and in the field of research which combines these two and which is the object of a rapidly growing research interest.

This focus on structural congruence and cognitive processing difficulty rather than on surface level incongruities, such as unexpected changes in prosody, timbre, tempo, and volume, are key to the applicability of this research to areas beyond music and language psychology. Indeed, this research identified similarities across domains in the processing of statistically improbable elements in structured stimuli which are processed more easily when they abide by implicitly learnt statistical-probability-based rules. In line with comments made by Kutas (1998), this approach therefore allowed a potential application of the findings from this research to the processing of other types of stimuli in which perception is also aided by implicitly learnt statistical probabilities. This in turn could grant insight into the basic emotional processes resulting from violated expectancies in different domains of human cognition.

9.2.3 Theoretically odd stimuli are perceived as odd, but a resolution is not always perceived as such

In the systematic review, concerns were raised over the fact that musical "oddness" was usually construed as a function of theoretical departure from the norm, with authors relying on the circle of fifths and chromaticism to create stimuli which were to sound unexpected (Patel et al., 1998; Koelsch, Gunter, et al., 2003). This method, at odds with the construction of incongruous stimuli in language (Van Berkum et al., 2005), left the question of whether theoretically odd elements were perceived as odd unaddressed.

The ratings obtained from participants with varying degrees of musical training did in fact show significantly higher oddity ratings for stimuli containing incongruities in both music and language (Chapter 4), despite participants repeatedly reporting difficulty deciding how odd a musical stimulus was in their post-experiment debrief. This supports the notion raised in the introductory chapters that mere exposure to stimuli presenting statistical structural regularities enables an implicit learning of rules which guide the processing of stimuli (Tillmann et al., 2000), even though participants may not have been aware of the specific elements causing them to perceive the stimuli as odd.

Indeed in *harmony*, non musicians showed no significant difference in ERP effects between the *normal* condition and the *incongruous-resolved* condition, either at the point of incongruity or at the point of resolution. These data suggest that oddity judgements of *incongruous-resolved* stimuli in *harmony* were based on the perception of the stimulus as a whole rather than on the detection of specific statistically improbable elements in this group of participants. In rhythm however, the ratings did seem to depend on the perception of specific elements within the stimulus. The oddity ratings for the *incongruous-resolved* stimuli were closer to the ratings for the *incongruous-unresolved* stimuli than expected. Inspection of the ERP data revealed a significant oddity-related ERP component at both the departure from the original beat pattern and the return to the original beat pattern. Despite the concerns expressed previously over the analysis of the rhythm data, these patterns do suggest oddity judgements based upon change detection in rhythm rather than upon overall harmonic integration in harmony.

These differences between *harmony* and *rhythm* suggest a difference in the time necessary for participants to settle into a new tonal or rhythmic framework. In *harmony*, since the shift back to the original key after the borrowed chord did not elicit an oddity-type response, we can assume that the original key still felt like the "home" key to the listener. In *rhythm* however, the return to the original beat did elicit an oddity-type component, which was in fact stronger than the effect observed at the point of incongruity. This suggests that participants had shifted their understanding of the beat structure away from the original beat and settled into the new beat by the time the resolution occurred. In both *harmony* and *rhythm*, the number of beats between the incongruity and its resolution was recorded for each stimulus. However, the analysis of ERP effects on different numbers of beats after the incongruity was not possible due to the limited number of trials with the same number of beats.

Considering the significant ERP findings in this experiment, future experiments using just one level of *hrss* from this stimulus set, and thus allowing time for more trials with the same number of beats between the incongruity and the resolution, would allow an investigation of the effect of incongruity duration on resolution patterns. Tracking each beat after the incongruity in a manner akin to the rhythm ERP studies reported by Münte et al. (2003) and Winkler, Karmos, and Näätänen (2002) would enable the investigation of the point at which the incongruous rhythmic pattern, or the accidentals from the key the incongruous chord is borrowed from, no longer elicit incongruity-type responses. Furthermore, assigning a different trigger code to the point of resolution based on the number of beats separating the point of resolution from the point of incongruity would enable the study of how many beats are necessary before the resolution is perceived as a new incongruity rather than a return to the "home" beat or key.

The behavioural data relating to the detection of oddity in the language stimuli showed a stronger distinction in ratings between the three levels of *congruence*. However, the ERP patterns for *semantics* showed no significant oddity-type response to the incongruity in musicians. The issue of differences between musicians and non musicians is discussed later in this chapter, with reference to the SSIRH. These data do however show that the lack of a response to the specific element which makes the sentence odd does not preclude an increased oddity rating in language also, suggesting a role for the overall integration of stimuli in oddity judgements in both music and language.

9.2.4 Participants reporting their "gut reaction" to music and language demonstrate overlapping labels between both domains

Certain aspects of participants' responses to stimuli showed a strong degree of overlap between music and language. Added to the fact that the strongest predictors of aesthetic ratings were stimulation ratings and tension ratings in both *harmony* and *semantics*, this observation supports the notion that similarities exist between the ways in which listeners respond to aesthetically-purposed incongruities in both music and language. Participants reported responses which overlapped with scales previously used to rate participants' aesthetic and affective responses to both music and language in the terms "exciting", "sad", "calm", and "peaceful" (Ali & Peynircioglu, 2006). The development of scales able to capture these facets of participants' responses provides a tool for future studies in this area of research. The categories which emerged from the analysis mapped onto previously used measures of grammaticality (Van Berkum et al., 2005), pleasantness, arousal (Koelsch, 2008) and tension (Steinbeis et al., 2006).

Considering similarities between theories of aesthetic in music, language, and other areas of human cognition in which incongruous elements have been shown to elicit emotions, such as visual art (Silvia, 2005), it would be interesting to see whether participants' responses to non auditory stimuli can be captured by similar labels. This could extend the methods developed in this research to the study of the role of incongruities in aesthetic experiences in other types of stimuli in which the input must also undergo structural organisation by the perceiver.

9.2.5 The notions of incongruity and resolution could account for the discrepancies in ERP components reported in seminal published studies in music ERP research

In *harmony*, the finding of a P600-type effect in musicians upon the incongruity and an N4/500-type effect at the point of "confirmation of nonsense" in non musicians elucidates part of the mystery of studies finding either a P600 (Patel et al., 1998; Patel, 2008) or an N400/500 (Hantz et al., 1997; Koelsch, Gunter, et al., 2003) in response to harmonically incongruous chords. Indeed, the parietal positive deflection found at the point of incongruity was similar to the Late Positive Component and the P600 reported by Besson et al. (1994), Miranda and Ullman (2007) and Patel et al. (1998), while the negative deflection in response to the harmonic elements which prevented harmonic closure showed a similar latency to the N5 reported by Jentschke et al. (2005), Miranda and Ullman (2007) and Steinbeis and Koelsch (2008b) in response to sequence-final incongruities which prevent harmonic closure. Its distribution (centro-parietal, right) was similar to the effect reported by Hantz et al. (1997) in its parietal distribution, and to the effect reported by Steinbeis and Koelsch (2008b). The discrepancies between stimuli, numbers of electrodes and experimental conditions call for caution in assigning a specific label to ERP components, as demonstrated in Tables 2.1 and 2.2. Nonetheless, it would appear that separating the notion of harmonic closure from the notion of harmonic congruence per se could be the key to progressing in our understanding of how the brain responds to harmonic incongruities.

Close inspection of researchers' musical stimuli also allows this distinction to explain the discrepancy between studies showing an interference of harmonic processing with semantics processing (Steinbeis & Koelsch, 2008b) and studies showing an interference of harmonic processing with syntactic processing (Slevc et al., 2009). The implications of this finding for both the interpretation of previous studies and the construction of stimuli in future studies are strong. It is hoped that this reconciliation of the disparate findings will help in the construction of the unified framework for research so desperately needed in this area (Juslin & Västfjäll, 2008; Vuust & Kringelbach, 2010).

9.2.6 Musical training affects participants' ERP and behavioural responses to both music and language

Considering the theoretical strength of the Shared Syntactic Integration Hypothesis (SSIRH), musicians and non musicians were compared throughout this thesis, in order to ascertain whether training in syntactic integration in one area (music) could affect the way in which integration takes place in a another area (language). Evidence of shared resources between music and language and interference of integrating semantics and syntax with integrating harmony (Steinbeis & Koelsch, 2008b; Slevc et al., 2009) made plausible the notion of integration transfer effects.

Musicians and non musicians did indeed show significant differences in their response to incongruities in both *harmony* and *semantics*. Both musicians and non musicians were recruited from a population of university-educated participants with similar experiences, speaking the same language and living in a society where the predominant music abides by the rules of Western tonal harmony. Adding to this the fact that the two musicianship samples in the experiment had a very similar mean age and gender split, musical training can safely be regarded as the factor underlying the difference between these groups. Possible interpretations and ideas for future studies building upon this result and the literature reported in the systematic review are provided in the final Section of this Chapter.

9.2.7 Overall: evidence for similarities in the role of incongruities in eliciting aesthetic and affective responses to music and language

9.2.7.1 Incongruities contribute to a higher aesthetic rating in harmony and semantics, possibly through the effect of tension and stimulation

In both *semantics* and *harmony*, a significant increase in aesthetic pleasantness ratings was seen when the stimuli contained an incongruity which later resolved. This finding provides timely support for theories of incongruity-based aesthetics using a theoretically driven, purpose-built and empirically validated set of normal, incongruous-resolved and incongruous-unresolved stimuli. The fact that stimuli in which the incongruity did not resolve were not judged as more aesthetically pleasing in all levels of *hrss* bar rhythm supports the previously neglected yet central role of resolution in these theories. Indeed the data presented in this experiment favour models describing the thrill derived from making sense of a temporarily incongruous stimulus (Vuust & Kringelbach, 2010) over those which suggest that novelty or complexity per se underlie aesthetic delight (e.g. Walker, 1973).

The two strongest predictors of aesthetic ratings were *stimulation* and *tension after listening*. Though the assumption of linear relationships between the scales is problematic in view of theories suggesting a bell-curve explanation for optimal arousal (Berlyne, 1963), the similarities seen in the two independent components in which an increase in aesthetic ratings was found as a consequence of a resolved incongruity suggest that stimulation and tension do play a role in eliciting aesthetic-affective responses.

When the analysis of ERP effects was added to the model, the similaries between responses to *harmony* and *semantics* were even more striking. In both these levels of *hrss*, an increased aesthetic rating in response to a resolved incongruity was associated with a significant or near significant increase in oddity ratings, no significant ERP effect linked to the incongruity, no significant effect of *congruence* in stimulation ratings, a larger increase in tension felt during listening between the *normal* and the *incongruous-resolved* condition than between the *incongruous-resolved* and the *incongruous-unresolved* condition, and no significant increase in the tension felt after listening to the stimulus. In the group showing no increase in aesthetic ratings, similar patterns were found in both levels of *hrss*.

9.2.7.2 On the interaction of the stimulus with the perceiving mind

The relationship between ERP components, oddity, stimulation and tension ratings emphasise the need to consider the interaction between the stimulus and the perceiving mind (Gaver & Mandler, 1987) in establishing a theory of how incongruities elicit aesthetic-affective responses. The behavioural data alone did not allow an understanding of why musicians and non musicians differed in their responses to resolved incongruities, since the data trends were identical in both groups in terms of oddity ratings. The fact that the group showing no ERP effect at the point of incongruity rated the incongruous-resolved stimulus as more asthetically pleasing emphasised the necessity of considering how the participant responds to the specific element constituting the aesthetically-purposed incongruity as a key step in any incongruity-based model of aesthetics. The literature surrounding aesthetic pleasantness has emphasised how the pleasure afforded by a specific piece of music varies widely from participant to participant (Blood & Zatorre, 2001). These data therefore call for caution with regards to the stimulus-response approach which seems to dominate the research in this area, and suggest as central a role for the perceiver's mind as for the incongruity itself in the eliciting of aesthetic-affective responses.

9.2.7.3 Music, language, incongruities... tension, stimulation, ERP effects... and aesthetics — shared patterns across domains

The data gathered in this research enabled the construction of the following tentative cross-domain model of the role of incongruities in eliciting aesthetic-affective responses to music and language:

1. A listener is exposed to a stimulus which he or she perceives to be slightly unusual but which can be integrated into a meaningful whole.

2. The listener's brain either responds or does not respond with a typical oddity pattern to the specific incongruous element.

    (a) If an ERP effect occurs, the stimulus is not perceived as more aesthetically pleasant than the normal stimulus. The data collected here suggest that this effect, reflecting processing difficulty, leads to levels of stimulation and tension beyond the levels required for a heightened aesthetic experience.

    (b) If no ERP effect occurs, either because the participant has not perceived the source of the incongruity, or because its integration requires no greater processing effort than a normal element, levels of stimulation and tension are heightened by the stimulus, but within the range leading to a heightened aesthetic experience.

3. The way in which the participant responds to the specific element at the source of the *incongruous-resolved* nature of the stimulus is key to the effect the incongruity has on the participants' aesthetic appreciation of the stimulus and affective response (tension or relaxation) felt as a result of listening to the stimulus.

9.3 Future directions

9.3.1 Room for improvement

9.3.1.1 Methods

The use of ERP methods in this research was motivated by the ability of this method to precisely determine the variations in electricity recorded at the scalp associated with a particular moment in time. This provided an ideal solution for building upon Meyer's statement that "a solid empirical basis for a theory of the arts [could] not be established until the biochemical functioning of the nervous system can be related to human cognitive and affective behaviour" (p. 34). By using ERP methods alongside behavioural methods, this research avoided the pitfall of merely speculating about the psychological implications of ERP effects and their link to aesthetic and affective responses, to create a data-rich approach to this question in need of more systematic investigation (Steinbeis et al., 2006; Thierry et al., 2008; Vuust & Kringelbach, 2010), as demonstrated by the disparity of effects to which the same label is assigned (see Tables 2.1 and 2.2) and the lack of significant progress in understanding the role of incongruities in eliciting aesthetic affective responses despite the methodological advances made over the last two decades.

The ERP method is however not without problems. One of the compromises in using ERPs is the inability to locate the area of the brain in which the electricity pattern observed at the scalp originates. Developments in MEG have enabled the localisation of specific effects such as the N500 in BA 21/37 (Koelsch et al., 2004) and the ERAN in BA 44 (Maess et al., 2001). Another compromise is its focus on the neocortex, the seat of perception and cognition, rather than on regions involved in emotions and reward (Blood & Zatorre, 2001), as could be speculated from theories of musical expectation (Huron, 2006; Vuust & Kringelbach, 2010). The application of PET scans, used to examine cerebral blood flow, to this area of research led to the discovery of the involvement of reward circuitry in the pleasure experienced during music chills (Blood & Zatorre, 2001). Though the study by Blood and Zatorre (2001) did not specify what characteristics of the music led to the chills described by the participants, a previous study of theirs demonstrated the involvement of paralimbic regions as a function of harmonic dissonance, linked to the perceived pleasantness or unpleasantness of musical stimuli (Blood et al., 1999). This suggests that methods allowing the investigations of limbic and paralimbic regions could shed further light on the neural mechanisms underlying the pleasure experienced during music and language listening.

Recent developments in the use of dopamine measurements via fMRI have enabled the discovery that musical anticipation which leads to musical pleasure is mediated by dopamine release (Gebauer & Vuust, 2010). Further research currently underway in the Music In the Brain Lab at Aarhus university aims to uncover further mechanisms linked to musical pleasure through the application of methods which allow the recording of levels of dopamine release during musical listening.

As in the case of ERP method development, it is crucial that a systematic and unified approach be adopted in order for theoretical progress to be made in this area of research. Nonetheless, the combination of methods allowing a precise time-locked examination of effects, an investigation of the involvement of areas known to underly emotions, and the measurements of dopamine release combine to provide exciting possibilities for future research on the role of incongruities in eliciting aesthetic and affective responses, and on structures underlying participants' experienced musical and linguistic pleasure.

9.3.1.2 Rhythm stimuli

The difficulties discussed in relation to the *rhythm* stimuli originate from the fact that in some stimuli it was not possible to maintain the original meter throughout the piece because of the insertion of a bar with an irregular number of beats. Future studies could use only the *rhythm* stimuli which do allow the original meter to be maintained and merely change to the emphasis of the beat (e.g. subtypes *a* and *b*). The maintaining of the original meter during rhythmically ambiguous rhythmic sections seems to be key to the tension and reward felt during the processing of rhythmic incongruities (Vuust et al., 2006). The methods used in this series of experiments did not allow the investigation of what meter participants adopted while listening to the stimuli. A separate set of studies using a tapping task would permit the analysis of the point at which participants switch meter, and would also permit the analysis of tension, confusion and aesthetic effects in contexts where participants are encouraged to hang onto the original meter of a piece. Such a study would allow the investigation of the effect of musical training on the ability to maintain an original meter and on the difficulty experienced in this task.

Using ERP methods, a trigger could also be placed on each beat after the incongruity to determine the point at which the new beat no longer sounds odd. After establishing the incongruity-type response to the onset of the rhythmically ambiguous section, this response could be followed through subsequent beats to see at which point it disappears. A different approach could be to create stimuli in which the duration of the ambiguous rhythmic section varies and analyse ERP effects upon resuming the original rhythmic pattern as a function of incongruity duration. Effects would also be analysed as a function of musical training to see whether training in the ability to maintain a meter results in differences in the point at which the original meter is lost leading to the resolution being perceived as a new incongruity.

9.3.2 Further research questions

9.3.2.1 What about expected incongruities?

Arising from any incongruity-based model is the question concerning repetitions. What happens when the listener knows not only that a certain statistically improbable musical or linguistic element will occur within the stimulus, but also exactly when that will take place?

Repeated exposure, leading to familiarity, removes the unexpected nature of the incongruity, which is central to the original incongruity-based models of aesthetics in music. Modular accounts of music processing, such as the one offered by Jackendoff (1991) to explain the continued enjoyment of repeated music, have proved controversial. Jackendoff's (1991) proposal states that the cognitive processes which react to unexpected elements in music are informationally encapsulated from the higher-level knowledge that the "unexpected" element is about to occur. This view has remained untested to date, since Meyer's (1956) original incongruity-based account of musical aesthetics has also lacked rigorous empirical investigation (Juslin & Västfjäll, 2008).

The same question could be asked of language stimuli. Would the repetition of incongruity-based metaphors diminish participants' aesthetic appreciation of the stimuli? The context in which sentences are presented has been shown to affect the amplitude of the ERP effects associated with low-cloze probability words. For instance, in a study by Nieuwland and Van Berkum (2006), presenting the semantically anomalous sentence "The peanut was *in love*" in the context of a story about peanut singing a song about his new girlfriend did not lead to a significant N400 effect on the words *in love* compared to the words *salted* in the sentence "The peanut was *salted*". However this is an issue of the context of meaningful semantic incongruities rather than an issue of repetition. Since the stimuli used in this research resulted in a similar pattern of effects for the processing of aesthetically-purposed incongruities in music and language, they provide the means of testing the effect of repetition in both domains.

A follow-up ERP experiment could be conducted using the stimulus set created and validated for future research in this thesis in a priming methodology borrowed from the study of sentence processing in psycholinguistics. Participants would listen to the stimuli presented in pairs: [normal - normal], [incongruous-resolved - incongruous-resolved] and [normal - incongruous-resolved]. Aesthetic ratings and ERP components would be compared separately in *harmony* and *semantics* in a 2 (iteration: first or second) by 2 (*congruence*: normal or incongruous-resolved) ANOVA. Incongruity-based models of aesthetics predict a significant main effect for *congruence*. Traditional understandings of memory and expectancies predict an interaction effect: if the unexpected nature of the incongruity is the source of the higher aesthetic rating, by repeating this element, it should be rendered less surprising, and thus, be less aesthetically pleasing. On the other hand, in the normal stimuli, since all elements are expected, no change in "unexpectness" would occur as a result of repetition, so there should be no modulation of ratings or ERP components beyond any general effects of repetition. Jackendoff's (1991) account however predicts no interaction effect, because of the modular cognitive architecture on which his theory is based.

A second analysis would examine the three levels of *congruence* of stimuli presented as the second member of the pair. The two *incongruous-resolved* stimuli (repeated and non-repeated) would be compared with the repeated *normal* baseline. If Jackendoff's (1991) account were correct, the two incongruous conditions should differ equally in comparison with the baseline. If however the two comparisons yielded a difference in magnitude, alternative explanations for the repeated enjoyment of familiar incongruities would need to be considered. This experiment would build upon the research carried out in this thesis to both benefit the development of research into the relationship between incongruities and aesthetics, and rekindle the debate surrounding Jackendoff's (1991) explanation of the repeated enjoyment of familiar incongruities.

9.3.2.2 What underlies the difference between patterns observed in musicians and non musicians in semantics?

A second question arising from the data presented in this thesis surrounds the differences observed between the responses of musicians and non musicians to the stimuli. The fact that musicians and non musicians responded differently to harmonic stimuli was not surprising. Such differences have been reported in studies investigating the processing of harmonic incongruities (Besson et al., 1994; Bonnel et al., 2001; Jentschke et al., 2005).

However, the differences observed in the processing of semantic incongruities were unexpected and call for more research. Theories such as the SSIRH and evidence suggesting shared mechanisms and even shared resources between music and language processing (Patel, 2003; Steinbeis & Koelsch, 2008b) makes sense of the fact that studies have reported beneficial effects of musical training on language processing (Forgeard et al., 2008; Overy, 2003; Schlaug et al., 2005; Tallal & Gaab, 2006). The data reported in Experiment 3 suggests that though musicians are equally able to indicate that a sentence containing an incongruity is more odd than a sentence containing no incongruity. While the oddity and confusion data trends from musicians were identical to those from non musicians, no ERP effects denoting semantic incongruity detection, or greater semantic processing effort, were found at the point of incongruity. A first explanation that musicians did not notice anything incongruous about the sentence is implausible because of the identical oddity rating trends observed between musicians and non musicians. A second explanation that musicians did not find the incongruity to be a source of increased processing difficulty is implausible because both musicians and non musicians reported significantly higher levels of confusion in the *incongruous-resolved* condition than in the *normal* condition.

A third explanation is that the point at which the incongruous nature of the sentence becomes apparent differs between groups. Previous evidence has pointed to the integration processes in music and language being shared (Patel, 2003) and to interferences of musical harmonic integration on linguistic integration (Slevc et al., 2009). Since musical training develops the ability to integrate elements according to specific rules or syntaxes, this development of a structure tracker or syntactic integrator should, according to these theories, lead to changes in the way in which integration takes place in language. Furthermore, considering the evidence for the beneficial effect of musical training on verbal memory (Ho et al., 2003) and verbal working memory span (Franklin et al., 2008), it could be that musicians allow a greater number of words to pass before integrating the sentence into a meaningful whole.

The timecourse of the integration of specific elements during sentence processing are typically investigated in psycholinguistics through self-paced reading (Koornneef &

Van Berkum, 2006) and eye-tracking (Featherstone & Sturt, 2010). To investigate the time-course of harmonic resolution, a self-paced listening task similar to the one used by Slevc et al. (2009) could be developed by associating each beat of the harmonic stimuli with a separate slide. However, of main importance here are the semantic stimuli. Self-paced reading could identify the points at which participants slow down in the processing of the sentences, an effect identified with increased processing difficulty (Van Berkum et al., 2005). Eye-tracking data could be used to see which points of the sentence participants return to in order to facilitate their comprehension of the sentence, by looking at both first-pass and second-pass reading times.

Such experiments would serve a double purpose. As a separate investigation from this thesis, they would allow further testing of the implications of the proposed SSIRH and provide insight into differences between musicians and non musicians in the integration of semantic information during sentence processing. The data collected for this thesis suggest that eye-tracking or self-paced reading studies would either show facilitated processing for semantic incongruities in musicians or a difference in the timecourse of semantic integration. The former would be evidenced by little delay upon the incongruity and subsequent words, while the latter would be evidenced by differences in the location of delays in the incremental processing of the sentences. Questions would then be raised as to the ways in which musical training contributes to different patterns in sentence integration, with potential explanations lying in the extensive memory training undergone by musicians who play by heart, the importance of processing musical elements as part of phrases, and the ability to listen to, and keep track of, information on a number of different levels which build up the texture of a piece as a whole. These experiments could also serve as a means of furthering the analysis carried out on the data collected during this thesis, by identifying, for each sentence, the point at which the semantic incongruity can be considered to be resolved. Since an ERP trigger was locked to each of the words following the incongruous word in the *semantic* stimuli, this information would allow the analysis of the ERP effects associated with the processing of the word providing the semantic resolution (or confirming the nonsensical nature of the sentence). This, in turn, could show further differences between musicians and non musicians, and would allow the addition of the patterns observed at the point of resolution to the cross-domain model of the relationship between incongruities and aesthetics emerging from the research presented in this thesis.

9.4 Concluding comments

The research carried out in this thesis demonstrated notable similarities between participants' reactions to aesthetically-purposed incongruities in music and language. In both harmony and semantics, an increased rating of aesthetic pleasantness was associated with a significantly increased oddity rating, a non-significant increase in stimulation rating, the absence of an ERP effect typically associated with oddity detection, and significantly higher ratings of tension felt during listening. Though initially promising a model which could explain incongruity-elicited aesthetic and affective responses to both music and language, the statistical analyses carried out were unable to paint a clear picture of the mechanism underlying these aesthetic-affective responses.

In some sense, it would have been somewhat of a tragedy if the complexity of aesthetic and affective responses to music and language had been boiled down to systematic patterns which could be used to create heightened aesthetic experiences on demand. The surprising nature of aesthetic delight could be seen as a force underlying cognitive development, which rewards the seeker of new experiences by providing a thrill when schema are developed to integrate elements which lie at the fringes of what our mental representations have grown to understand thus far. In that light, surely the beauty here, for the academic researcher as well as the musico- and lingua- phile, lies in the unexpected nature of some of the data trends which prompt further research, the regularities which hint at other carefully structured shared mechanism across domains yet to be uncovered, and the elusive nature of the relationship between incongruities, the brain's reaction and our enjoyment of what we see, read and hear.

References

Aiken, H. (1951). The aesthetic relevance of belief. *The Journal of Aesthetics and Art Criticism*, *9*(4), 301-315.

Ali, S., & Peynircioglu, Z. F. (2006). Songs and emotions: Are lyrics and melodies equal partners? *Psychology of Music*, *34*(4), 511-534.

Balkwill, L.-L., & Forde Thompson, W. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*(1), 43-64.

Bartolo, A., Benuzzi, F., Nocetti, L., Beraldi, P., & Nichelli, P. (2006). Humour comprehension and appreciation: An fmri study. *Journal of Cognitive Neuroscience*, *18*(11), 1789-1798.

Beisteiner, R., Erdler, M., Mayer, D., Gartus, A., Edward, V., Kaindl, T., et al. (1999). A marker for differentiation of capabilities for processing of musical harmonies as detected by magnetoencephalography in musicians. *Neuroscience Letters*, *277*, 37-40.

Bella, S. D., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics*, *65*(7), 1019-1028.

Berlyne, D. (1963). Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings. *Canadian Journal of Psychology*, *17*, 274-290.

Besson, M. (1998). Meaning, structure, and time in language and music. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *17*(4-5), 921-950.

Besson, M. (1999). The musical brain: Neural substrates of music perception. *Journal of New Music Research*, *28*(3), 246-256.

Besson, M., Faita, F., & Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non musicians. *Neuroscience Letters*, *168*(1-2), 101-105.

Besson, M., & Friederici, A. (2005). Language and music - a comparison introduction.

In *Neurosciences and music ii: From perception to performance* (Vol. 1060, p. 57-58). New York: New York Acad Sciences.

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, *24*(1), 14-25.

Besson, M., & Schon, D. (2001). Comparison between language and music. In *Biological foundations of music* (Vol. 930, p. 232-258). New York: New York Acad Sciences.

Blasko, D. G., & Kazmerski, V. A. (2006). Erp correlates of individual differences in the comprehension of nonliteral language. *Metaphor and Symbol*, *21*(4), 267-284.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain refions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, *98*(20), 11818 - 11823.

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience*, *2*(4), 383 - 387.

Bod, R. (2001). Memory based model for music analysis: Challending the gestalt principles. *Journal of New Music Research*, *31*(1), 26-36.

Bod, R. (2002). A unified model of structural organization in language and music. *Journal of Artificial Intelligence Research*, *17*, 289-308.

Bonnaud, V., Gil, R., & Ingrand, P. (2002). Metaphorical and non-metaphorical links: a behavioral and erp study in young and elderly adults. *Neurophysiologie Clinique*, *32*, 258-268.

Bonnel, A.-M., Faita, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics*, *63*(7), 1201-1213.

Bowdle, B., & Gentner, D. (2005). The career of metaphor. *Psychological Review*, *112*(1), 193-216.

Bradshaw, J. L., Nettleton, & Geffen, G. (1971). Ear differences and delayed auditory feedback - effects on a speech and a music task. *Journal of Experimental*

Psychology, *91*(1), 85-86.

Brattico, E., & Jacobsen, T. (2009). Subjective appraisal of music: Neuroimaging evidence. *Annals of the New York Academy of Sciences, The Neurosciences and Music III: Disorders and Plasicity*, *1169*, 308-317.

Bregman, A. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

Broadbent, M. H. P., D.E. Broadbent. (1981). Recency effects in visual memory. *Quarterly Journal of Experimental Psychology*, *33*(A), 1-15.

Brown, S. (2001). Are music and language homologues? In *Biological foundations of music* (Vol. 930, p. 372-374). New York: New York Acad Sciences.

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: A pet study of the generation of melodies and sentences. *European Journal of Neuroscience*, *23*(10), 2791-2803.

Budd, M. (1985). *Music and the emotions: The philosophical theories.* Boston: Routledge & Kegan Paul.

Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., et al. (2006). Song and speech: Brain regions involved with perception and covert production. *Neuroimage*, *31*(3), 1327-1342.

Celesia, G. G. (1976). Organization of auditory cortical areas in man. *Brain*, *99*(3), 403-414.

Cheatham, B. H. (1990). Speech and music draw attention to the right ear in stereophonic listening. *Perceptual and Motor Skills*, *70*(1), 227-231.

Chiappe, P., & Schmuckler, M. A. (1997). Phrasing influences the recognition of melodies. *Psychonomic Bulletin & Review*, *4*(2), 254-259.

Chomsky, N., & Halle, M. (1968). *The sound pattern of english.* London: Harper & Row.

Clark, E., & Clark, H. (1979). When nouns surface as verbs. *Language*, *55*(4), 767-811.

Coffman, D. D., Gfeller, K., & Eckert, M. (1995). Effect of textual setting, training, and gender on emotional response to verbal and musical information. *Psychomu-*

sicology, *14*(1), 117-136.

Collard, R., Vos, P., & Leewenberg, E. (1981). What melody tells about metre in music. *Zeitschrift für Psychologie*, *189*, 25-33.

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958-968.

Coulson, S., & Van Petten, C. (2007). A special role for the right hemisphere in metaphor comprehension? erp evidence from hemifield presentation. *Brain Research*, *1146*, 128-145.

Coulson, S., & Williams, R. F. (2005). Hemispheric asymmetries and joke comprehension. *Neuropsychologia*, *43*(1), 128-141.

Craik, F. I. M., & Lockhart, R. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, *11*, 671-684.

Cunningham, W. A., & Derks, P. (2005). Humor appreciation and latency of comprehension. *Humor-International Journal of Humour Research*, *18*(4), 389-403.

Daltrozzo, J., & Schön, D. (2009a). Conceptual processing in music as revealed by n400 effects on words and musical targets. *Journal of Cognitive Neuroscience*, *21*(10), 1882-1892.

Daltrozzo, J., & Schön, D. (2009b). Is conceptual processing in music automatic? an electrophysiological approach. *Brain Research*, *1270*, 88-94.

Deliege, I. (1996). Cue abstraction as a component of categorisation processes iin music listening. *Psychology of Music*, *24*(2), 131-156.

DePascalis, V., Marucci, F. S., Penna, M. P., Labbrozzi, D., & al. et. (1987). Event-related potentials in musically sophisticated and unsophisticated subjects: A study on hemispheric specialization. *Neuropsychologia*, *25*(6), 947-955.

Dowling, W., & Harwood, D. L. (1986). *Music cognition*. Orlando: Academic Press.

Ducreux, D., Marsot-Dupuch, K., Lasjaunias, P., Oppenheim, C., & Fredy, D. (2003). Lyrical and musical auditive mental imagery in functional mri. *Journal of Neuroradiology*, *30*(1), 18-24.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flecible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Featherstone, C. R., & Sturt, P. (2010). Because there was a cause for concern: an investigation into a word-specific predition account of the implicit-causality effect. *Quarterly Journal of Experimental Psychology*, *63*(1), 3-15.

Flores-Gutierrez, E. O., Diaz, J.-L., Barrios, F. A., Favila-Humara, R., Guevara, M. A., Rio-Portilla, Y. del, et al. (2007). Metabolic and electric brain patterns during pleasant and unpleasant emotions induced by music masterpieces. *International Journal of Psychophysiology*, *65*(1), 69-84.

Fodor, J. A., & Bever, T. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behaviour*, *4*, 414-420.

Forgeard, M., Winner, E., Norton, A., & Schlaug, G. (2008). Practicing a musical instrument in childhoosd is associated ith enhanced verbal ability and nonverbal reasoning. *PloS One*, *3*(10), e3566.

Franklin, M. S., Moore, K. S., Yip, C.-Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*, 353-365.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies.* Unversity of Connecticut: PhD Thesis.

Friston, K. (2005). A theory of cortincal responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 815-836.

Gaillard, A. W. K. (1988). Problems and paradigms in erp research. *Biological Psychology*, *26*, 91-109.

Galinanes, C. L. (2005). Funny fiction; or, jokes and their relation to the humourous novel. *Poetics Today*, *26*(1), 76-111.

Gardner, B. (1965). *Up to the line to death.* London: Methuen.

Gaver, W. W., & Mandler, G. (1987). Play it again, sam: On liking music. *Cognition* *and Emotion*, *1*, 259-282.

Gebauer, L., & Vuust, P. (2010). The rewards of music listening: The role of midbrain dopamine in musical anticipation. *Proceedings of the 11th International Conference on Music Percerption and Cognition, Seattle.*

Geisler, W., Perry, J., Super, B., & Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*(6), 711-724.

Gfeller, K., Asmus, E. P., & Eckert, M. (1991). An investigation of emotional response to music and text. *Psychology of Music*, *19*(2), 128-141.

Gibson, E. (1998). Linguisticcomplexity:localityofsyntacticdependencies. *Cognition*, *68*, 1-76.

Goodglass, H., & Calderon, M. (1977). Parallel processing of verbal and musical stimuli in right and left hemispheres. *Neuropsychologia*, *15*(3), 397-407.

Greene, R. L., & Samuel, A. G. (1986). Recency and suffix effects in serial recall of musical stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(4), 517-524.

Gross, D. (2009). Still relevant after decades, the beatles set to rock. *http://edition.cnn.com/2009/SHOWBIZ/Music/09/04/beatles.999/index.html*, Article retrived online 26th September 2010.

Groussard, M., Viader, F., Hubert, V., Landeau, B., Abbas, A., Desgranges, B., et al. (2010). Musical and verbal semantic memory: Two distinct neural networks? *NeuroImage*, *49*, 2764-2773.

Grout, D. (1973). *A history of western music.* London: Dent.

Hagoort, P. (2005). On broca, brain, and binding: a new framework. *Trends in cognitive science*, *9*, 416-423.

Hagoort, P., & Brown, C. M. (2000). Erp effects of listening to speech compared to reading: The p600/sps to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *11*(1), 1531-1549.

Hale, J. (2001). A probabilistic early parser as a psycholinguistic model. In *North* *american chapter of the association for computational linguistics - second meeting of the north american chapter of the association for computational linguistics on language technologies 2001* (p. 1-8). Pittsburgh, Pennsylvania: Association for Computational Linguistics.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643-672.

Hantz, E. C., Kreilick, K. G., Kananen, W., & Swartz, K. P. (1997). Neural responses to melodic and harmonic closure: An event-related-potential study. *Music Perception*, *15*(1), 69-98.

Harley, T. A. (2001). *The psychology of language: from data to theory*. Hove: Psychology press.

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Musical training improves verbal but not visual memory: Cross sectional and longitudinal explorations in children. *Neuropsychology*, *17*(3), 439-450.

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

Isel, F., Alter, K., & Friederici, A. D. (2005). Influence of prosodic information on the processing of split particles: ERP evidence from spoken German. *Journal of Cognitive Neuroscience*, *17*(1), 154-167.

Jackendoff, R. (1980). *A deep parallel between music and language*. Bloomington, Ind.: Indiana University Linguistics Club.

Jackendoff, R. (1991). Musical parsing and musical affect. *Music Perception*, *9*(2), 199-230.

Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Journal of Experimental Psychology*, *41*(4), 291-297.

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the develpment of syntax processing in children. *NeuroImage*, *47*, 735-744.

Jentschke, S., Koelsch, S., & Friederici, A. D. (2005). Investigating the relationship of music and language in children: Influences of musical training and language impairment. In G. Avanzini, L. Lopez, S. Koelsch, & M. Manjno (Eds.), *The neurosciences and music ii: From perception to performance* (p. 231-242). New York, NY: New York Academy of Sciences.

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of musix-syntactic processing. *Journal of Cognitive Neuroscience*, *20*(11), 1940-1951.

Jentzch, I., & Sommer, W. (2001). Sequence-sensitive subcomponents of p300: Topographical analyses and dipole source localization. *Psychophysiology*, *38*, 607 - 621.

Jones, M. R. (1987). Dynamic pattern strucgure in music: Recent theory and research. *Perception & Psychophysics*, *41*, 621-624.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*, 313-319.

Jongsma, M., Desain, P., & Honing, H. (2004). Rhythmic context influences the auditory evoked potentials of musicians and nonmusicians. *Biological Psychology*, *66*, 129-152.

Jurasfky, D. (2003). Probabilistic modelling in psycholinguistics: linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics.* Cambridge, MA: MIT Press.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioural and Brain Sciences*, *31*, 559-621.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. J. (2000). The p600 as an index of syntactic integration difficulty. *Language & Cognitive Processes*, *15*, 159-201.

Katz, A. (2005). Preface. In H. Colston & A. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (p. ix-xi). London: Lawrence Erlbaum Associates.

Kazmerski, V., Blasko, D., & Dessalegn, B. (2003). Erp and behavioral evidence of individual differences in metaphor comprehension. *Memory & Cognition*, *31*, 673-689.

Kerkhofs, R., Vonk, W., Schriefers, H., & Chwilla, D. J. (2008). Sentence processing in the visual and auditory modality: Do comma and prosodic break have parallel functions? *Brain Research*, *1224*, 102-118.

Kimura, D. (1964). Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology*, *16*(4), 355-358.

Kluender, R., & Kutas, M. (1993). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, *8*(4), 573-633.

Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, *18*(5), 1169-1178.

Koelsch, S., & Friederici, A. D. (2003). Toward the neural basis of processing structure in music - comparative results of different neurophysiological investigation methods. In *Neurosciences and music* (p. 15-28). New York: New York Academy of Sciences.

Koelsch, S., Grossmann, T., Gunter, T. C., Hahne, A., Schroger, E., & Friederici, A. D. (2003). Children processing music: Electric brain responses reveal musical competence and gender differences. *Journal of Cognitive Neuroscience*, *15*(5), 683-693.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "non musicians" are musical. *Journal of Cognitive Neuroscience*, *12*, 520-541.

Koelsch, S., Gunter, T., Schroger, E., & Friederici, A. D. (2003). Processing tonal modulations: An erp study. *Journal of Cognitive Neuroscience*, *15*(8), 1149-1159.

Koelsch, S., Gunter, T. C., Cramon, D. Y. von, Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *Neuroimage*, *17*(2), 956-966.

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An erp study. *Journal of Cognitive Neuroscience*, *17*(10), 1565-1577.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302-307.

Koelsch, S., Kilches, S., Steinbeis, N., & Schelinski, S. (2008). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLos One*, *3*(7), e2631.

Koelsch, S., Maess, B., Grossmann, T., & Friederici, A. D. (2003). Electric brain responses reveal gender differences in music processing. *Neuroreport: For Rapid Communication of Neuroscience Research*, *14*(5), 709-713.

Koelsch, S., Maess, B., Gunter, T. C., & Friederici, A. D. (2001). Neapolitan chords activate the area of broca - a magnetoencephalographic study. In *Biological foundations of music* (Vol. 930, p. 420-421). New York: New York Academy of Sciences.

Koelsch, S., & Mulder, J. (2002). Electric brain responses to inappropriate harmonies during listening to expressive music. *Clinical Neuropsychology*, *113*, 862-869.

Koelsch, S., Schroger, E., & Gunter, T. (2002). Music matters: Preattetive musicality of the human brain. *Psychophysiology*, *39*, 38-48.

Koelsch, S., Schroger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport: For Rapid Communication of Neuroscience Research*, *10*(6), 1309-1313.

Koffka, K. (1935). *Principles of gestalt psychology.* New York: Harcourt, Brace & Co.

Konovalov, V., & Otmakhova, N. (1984). Eeg manifestations of functional asymmetry of the human cerebral cortex during perception of words and music. *Human Physiology*, *9*(4), 250-255.

Koornneef, A. W., & Van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *34*, 543-566.

Krumhansl, C. L. (1985). Perceiving tonal structure in music. *American Scientist*, *73*, 371-378.

Krumhansl, C. L. (2002). Music: a link between cognition and emotion. *Current Directions in Psychological Science*, *11*(2), 45-50.

Kutas, M. (1998). Current thinking on language structures. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *17*(4-5), 951-969.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*, 203-205.

Kutas, M., & Hillyard, S. A. (1982). The lateral distribution of event-related potentials during sentence processing. *Neuropsychologia*, *20*(5), 579-590.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. *Advances in Psychophysiology*, *3*, 139-187.

LaBarba, R. C., & Kingsberg, S. A. (1990). Cerebral lateralization of familiar and unfamiliar music perception in nonmusicians: A dual task approach. *Cortex*, *26*(4), 567-574.

Lau, E. F., Phillips, C., & Peoppel, D. (2008). A cortical network for semantics: (de)constructing the n400. *Nature Reviews Neuroscience*, *9*, 920-933.

Lee, C.-L., & Federmeier, K. (2006). To mind the mind: An event-related potential study of word class and semantic ambiguity. *Brain Research*, *1081*, 191-202.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: MIT Press.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: a possible role for brodmann area 47 in temporal coherence. *Neuroimage*, *20*(4), 2142-2152.

Levitin, D. J., & Menon, V. (2005). The neural locus of temporal structure and expectancies in music: Evidence from functional neuroimaging at 3 tesla. *Music Perception*, *22*(3), 563-575.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *22*(3), 1126-1177.

Lim, C. J. (2006). Structural and functional neural correlates of music perception. *Anatomical Record Part a-Discoveries in Molecular Cellular and Evolutionary Biology*, *288A*(4), 435-446.

Lindstrom, E. (2006). Impact of melodic organization on perceived structure and emotional expression in music. *Musicae Scientiae*, *10*(1), 85-117.

Longuet-Higgins, H. C. (1994). Artificial-intelligence and musical cognition. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences*, *349*(1689), 103-113.

Lovelock, W. (1957). *The rudiments of music*. London: Bell & Hyman.

Luck, S. J. (2006). *An introduction to the event-related potential technique*. London: MIT Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of szntactic ambiguity resolution. *Psychological Review*, *101*, 676-703.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in broca's area: An meg study. *Nature Neuroscience*, *4*(5), 540-545.

Magne, C., Mistuko, A., Astesano, C., Gordon, R. L., Ystad, S., Farner, S., et al. (2005). Comparison of rhythmic processing in language and music. *Journal of Music and Meaning*, *3*(5), sec.5.4.3.

Mandler, G. (n.d.). The structure of value. In M. S. Clark & S. T. Fiske (Eds.), *Affect and cognition* (p. 3-36). Hillsdale, NJ: Erlbaum.

Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. London: MIT Press.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1-71.

Martin-Loeches, M., Casado, P., Gonzalo, R., De Heras, L., & Fernandez-Frias, C. (2006). Brain potentials to mathematical syntax problems. *Psychophysiology*, *43*(6), 579-591.

McLane, A. (1996). Music as information. *Annual Review of Information Science and Technology*, *31*, 225-262.

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, *21*(3), 289-311.

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

Meyer, L. B. (1957). Meaning in music and information theory. *The Journal of Aesthetics and Art Criticism*, *15*(4), 412-424.

Meyer, L. B. (1974). Concerning the sciences, the arts and the humanities. *Critical Inquiry*, *1*(4), 163-217.

Meyer, L. B. (2000). *The spheres of music: a gathering of essays*. Chicago: University of Chicago Press.

Meyer, L. B., & Schvaneveldt, R. (1971). Faciliatation in recognizing pairs of words: Evidence of a dependence between retreieval operations. *Journal of Experimental Psychology*, *90*, 227-234.

Miranda, R. A., & Ullman, M. T. (2007). Double dissociation between rules and memory in music: An event-related potential study. *Neuroimage*, *38*(2), 331-345.

Mirz, F., Ovesen, T., Ishizu, K., Johannsen, P., Madsen, S., Gjedde, A., et al. (1999). Stimulus-dependent central processing of auditory stimuli: A pet study. *Scandinavian Audiology*, *28*(3), 161-169.

Müller, M., Höfel, L., Brattico, E., & Jacobsen, T. (2010). Aesthetic judgements of music in experts and laypersons: An erp study. *International Journal of Psychophysiology*, *76*, 40-51.

Münte, T. F., Nager, W., Beiss, T., Shroeder, C., & Altenmüller, E. (2003). Specialization of the specialized: Electrophysiological investigations in professional musicians. *Annals New York Academy of Sciences*, *999*, 131-139.

Naatanen, R. (1976). The perception of speech sounds by the human brain as reflected by the mismatch negativity (mmn) and its magnetic equivalent (mmnm). *Brain*, *99*(3), 403-414.

Nan, Y., Knosche, T. R., & Friederici, A. D. (2006). The perception of musical phrase structure: A cross-cultural ERP study. *Brain Research*, *1094*, 179-191.

Neuhaus, C., Knosche, T. R., & Friederici, A. D. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *Journal of Cognitive Neuroscience*, *18*(3), 472-493.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of COgnitive Neuroscience*, *18*(7), 1098-1111.

Norrick, N. R. (2003). Issues in conversational joking. *Journal of Pragmatics*, *35*(9), 1333-1359.

Noveck, I., Bianco, M., & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol*, *16*(1-2), 109-121.

Ortony, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory*, *25*(1), 45-53.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions.* Cambridge: Cambridge University Press.

Osterhout, L., Allen, M. D., McLaughlin, J., & Inoue, K. (2002). Brain potentials elicited by prose-embedded linguistic anomalies. *Memory & Cognition*, *30*(8), 1304-1312.

Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806.

Overy, K. (2003). Dyslexia and music: From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*, *999*, 497-505.

Paller, K. A., McCarthy, G., & Wood, C. C. (1992). Event-related potentials elicited by deviant endings to melodies. *Psychophysiology*, *29*(2), 202-206.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *17*(3), 407-421.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674-681.

Patel, A. D. (2005). The relationship of music to the melody of speech and to syntactic processing disorders in aphasia. [references]. In G. Avanzini, L. Lopez, S. Koelsch, & M. Manjno (Eds.), *The neurosciences and music ii: From perception to performance* (p. 59-70). New York, NY: New York Academy of Sciences. (xii 490 Chapter)

Patel, A. D. (2007). Language, music, and the brain: a resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems.* Oxford: Oxford University Press.

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

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717-733.

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). Functional congribution of the basal ganglia to emotional prosody: Evidence from erps. *Brain Research*, *1217*, 171-178.

Peretz, I., Radeau, M., & Arguin, M. (2004). Two-way interactions between music and language: Evidence from priming recognition of tune and lyrics in familiar songs. *Memory & Cognition*, *32*(1), 142-152.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). Erp effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*, 407-428.

Piston, W. (1978). *Harmony.* London: Gollancz.

Pomerantz, J. R. (1981). Perceptual organization in information processing. In

M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (p. 141-180). New Jersey: Lawrence Erlbaum Associates.

Ponsford, D., Wiggins, G., & Mellish, C. (1999). Statistical learning of harmonic movement. *Journal of New Music Research*, *28*(2), 150-177.

Poulin-Charronnat, B., Bigand, E., Madurall, F., & Peereman, R. (2006). Musical structure modulates semantic priming in vocal music. *Cognition*, *94*, B67-B78.

Pynte, J., Besson, M., Robichon, F., & Poli, J. (1996). The time course of metaphor comprehension: An event-related potential study. *Brain and Language*, *55*, 293-316.

Raffman, D. (1993). *Language, music, and mind.* Cambridge, MA.: MIT Press.

Raphael, C. (1999). Automatic segmentation of acoustic musical signals using hidden markov models. *IEEE Transactions on Pattern Analzsis and Machine Intelligence*, *21*(4), 360-370.

Regnault, P., & Besson, M. (1998). Attentional modulation of language and music processing in opera: an electrophysiological approach. *Cahiers De Psychologie Cognitive-Current Psychology of Cognition*, *17*(4-5), 1069-1070.

Ritchie, D. (2005). Frame-shifting in humour and irony. *Metaphor and Symbol*, *20*(4), 275-294.

Rivenez, M., Guillaume, A., & Darwin, C. (2006). Processing unattended speech. *The journal of the acoustical society of america*, *119*, 1027-4040.

Roberts, L. A. (1986). Modality and suffix effects in memory for melodic and harmonic musical materials. *Cognitive Psychology*, *18*(2), 123-157.

Rock, I., & Mack, A. (1994). Attention and perceptual organization. In S. Ballesteros (Ed.), *Cognitive approaches to human perception* (p. 23-41). New Jersey: Erlbaum.

Rock, I., & Palmer, S. (1990). The legacy of gestalt psychology. *Scientific American*, *263*(6), 84-90.

Ross, E. D. (2010). Cerebral localization of functions and the neurology of language:

Fact verus fiction or is it something else? *The Neuroscientist*, *16*(3), 222-243.

Rugg, M., Mark, R. E., Gilchrist, J., & Roberts, R. (1997). Erp repetition effects in indirect and direct tasks: Effects of age and interitem lag. *Psychophysiology*, *34*, 572-586.

Saffran, J. R. (2003). Musical learning and language development. In *Neurosciences and music* (Vol. 999, p. 397-401).

Saito, S. (2001). The phonological loop and memory for rhythms: An individual differences approach. *Memory*, *9*(4-6), 313-322.

Saito, S., & Ishio, A. (1998). Rhythmic information in working memory: Effects of concurrent articulation on reproduction of rhythms. *Japanese Psychological Research*, *40*(1), 10-18.

Sammler, D., Baird, A., Valabrgue, R., S., C., Dupont, S., & Belin, P. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *The Journal of Neuroscience*, *30*(10), 3572-3578.

Schendel, Z. A., & Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory & Cognition*, *35*(4), 640-650.

Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences*, *1060*, 219-230.

Schmithorst, V. J. (2005). Separate cortical networks involved in music perception: preliminary functional mri evidence for modularity of music processing. *Neuroimage*, *25*(2), 444-451.

Schmuckler, M. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception*, *14*, 109-150.

Schulkind, M. D., Posner, R. J., & Rubin, D. C. (2003). Musical features that facilitate melody identification: How do you know it's "your" song when they finally play it? *Music Perception*, *21*(2), 217-249.

Segalowitz, S., & Plantery, P. (1985). Music draws attention to the left and speech draws attention to the right. *Brain and Cognition*, *4*(1), 1-6.

Sekuler, R., & Blake, R. (2002). *Perception.* New York: McGraw Hill.

Silvia, P. J. (2005). Emotional responses to art: From collation and arousal to cognition and emotion. *Review of General Psychology*, *9*, 342-357.

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistic musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic BUlletin & Review*, *16*(2), 374-381.

Sloboda, J. A., & Gregory, A. H. (1980). The psychological reality of musical segments. *Canadian Journal of Psychology*, *34*, 274-280.

Smyth, M. M., Hay, D. C., Hitch, G. J., & Horton, N. J. (2005). Serial position memory in the visual-spacial domain: Reconstructing sequences of unfamiliar faces. *Quarterly Journal of Experimental Psychology*, *58A*(5), 909-930.

Spillmann, L. (2006). Phenomenology and neurophysiological correlations: Two approaches to perception research. *Vision Research*, *49*(12), 1507-1521.

Squire, L., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380-1386.

Steinbeis, N., & Koelsch, S. (2008a). Comparing the processing of music and language meaning using eeg and fmri provides evidence for similar and distinct neural representations. *PLoS One*, *3*(5), e2226.

Steinbeis, N., & Koelsch, S. (2008b). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, *18*(5), 1169-1178.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, *18*(8), 1380-1393.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, *86*(1), 142-164.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*(2), 191-196.

Stratton, V. N., & Zalanowski, A. H. (1994). Affective impact of music vs. lyrics. *Empirical Studies of the Arts*, *12*(2), 173-184.

Sundberg, J., & Lindblom, B. (1976). Generative theories in language and music descriptions. *Cognition*, *4*(1), 99-122.

Sussman, E. (2005). Integration and segregation in auditory scene analysis. *The Journal of the Acoustical Society of America*, *117*, 1285-1298.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potential. *Science*, *150*, 1436-1439.

Swinney, D., & Love, T. (1998). The processing of discontinuous dependencies in language and music. *Music Perception*, *16*(1), 63-78.

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences*, *29*(7), 382-390.

Tartter, V. C., Gomes, H., Dubrovsky, B., Molholm, S., & Stewart, R. V. (2002). Novel metaphors appear anomalous at least momentarily: Evidence from n400. *Brain and Language*, *80*(3), 488-509.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*, 415.

Tervaniemi, M., Szameitat, A. J., Kruck, S., Schroger, E., Alter, K., De Baene, W., et al. (2006). From air oscillations to music and speech: Functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. *Journal of Neuroscience*, *26*(34), 8647-8652.

Tettamanti, M., & Weniger, D. (2006). Broca's area: A supramodal hierarchical processor? *Cortex*, *42*(4), 491-494.

Thierry, G., Martin, C., Gonzalez-Diaz, V., Rezaie, R., Roberts, N., & Davis, M. (2008). Event-related potential characterisation of the shakespearean functional shift in narrative sentence structure. *Neuroimage*, *40*, 923-931.

Tillmann, B., Bharucha, J., & Bigand, E. (2000). Implicit learning of tonality: a self-organising approach. *Psychology review*, *107*, 885-913.

Ullman, M. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews Neuroscience*, *2*, 717-726.

Uttal, W. (2001). *The new phrenology: The limits of localizing cognitive processes in the brain*. Cambridge, Mass.: London: MIT Press.

Van Berkum, J. J. A., Brown, C., Zwitserlood, P., Koojman, V., & Hagoort, P. (2005). Anticipating upcoming words in discource: evidence from erps and reading times. *Journal of Experimental Psychology-Learning Memory and Cognition*, *31*(3), 443-467.

Van Berkum, J. J. A., Hagoort, P., & Brown, M., Colin. (1999). Semantic integration in sentences and discourse: Evidence from the n400. *Journal of Cognitive Neuroscience*, *11*(6), 657-671.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, memory and cognition*, *25*(2), 394-417.

Vecera, S., & Farah, M. (1997). Is visual image segmentation a bottom-up or an interactive process? *Perception and Psychophysics*, *59*(8), 1280-1296.

Vuust, P., & Kringelbach, M. L. (2010). The pleasure of making sense of music. *Interdisciplinary science reviews*, *35*(2), 166-182.

Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Ostergaard, L. (2006). It don't mean a thing ... keeping the rhythm during polyrhythmic tension, activates language areas (ba47). *Neuroimage*, *31*(2), 832-841.

Walker, E. L. (1973). Psychological complexity and preference: A hedgehog theory of behaviour. In D. E. Berlyne & K. B. Madsen (Eds.), *Pleasure, reward, preference.* New York: Academy Press.

Wallentin, M., Nielson, A. H., Friis-Olivaruius, M., Vuust, C., & Vuust, P. (2010). The musical ear teat, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*, 188-196.

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory & Cognition*, *38*(2), 163-175.

Winkler, I., Karmos, G., & Näätänen, R. (2002). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, *742*, 239-252.

Ystad, S., Magne, C., Pallone, G., Aramaki, M., Besson, M., & Kronland-Martinet, R. (2007). Electrophysiological study of algorithmically processed metric/rhythmic variations in language and music. *Journal on Audio, Speech and Music Processing*, *14*(1), 188-196.

Yus, F. (2003). Humor and the search for relevance. *Journal of Pragmatics*, *35*(9), 1295-1331.

Appendix A

Initial observations of equivalences between music and language

distance dependencies
lish, 1999)

g a hierarchical structure
ds/intervals), tonal
temporal clock (rhythm)

ts (Bod, 2002)
her traditional
l violation of its own
and aesthetic purposes

d Levitin and Menon
synt of music

pirical test (Patel, 2003)

c might relate to
herent units of musical
t al., 2006)
ing are in fact theories of
temporal coherence

red mostly self-
on, 2001)
empo and volume

m of analytic approaches
al and/or cultural
al/internal world, and
n & Schon, 2001)

mbiguities, redundancies
os (appearance of score /
cLane, 1996)

998)
s (Patel, 1998) so best to

1998)
ructural or aesthetic

hes (Besson & Schon,

	Language	Music
Hierarchical structure	▪ Chomsky and Halle's (1968) generative phonology of English to establish where stress goes	▪ Idem (Sundberg & Lindblom, 1976), (Lerdahl and Jackendoff, 1980, 1983)
Perception	▪ Facilitated by rules (Sundberg & Lindblom, 1976) ▪ The perceptual input undergoes a process of hierarchical structuring which is not found in the input itself (Bod, 2002)	
Importance of notation	▪ In poetry, stanzas and lines reveal grouping (Longuethiggins, 1994)	▪ In music, notation reveals tonality and rhythmic grouping (Longuethiggins, 1994)
Parsing / structuring of sound	▪ Importance of statistical preference (corpus) (Hale, 2001)	▪ Statistical (corpus) (Sundberg & Lindblom, 1976) ▪ RHYTHM: Once the meter is known and the barlines correctly placed, each note can immediately be assigned a place in the rhythmic hierarchy. Can be revised (i.e. if faced with persistent syncopation) (Longuethiggins, 1994) ▪ TONALITY: once a key has been suggested by the first few notes, the following notes virtually fall into place. Can be revised (if resulting pattern of relations between pitches suggests chance of key) (Longuethiggins, 1994)
	▪ Likelihood principle (Manning & Schultze, 1999), most appropriate seen as most likely	▪ Likelihood principle (Raphael, 1999; Bod, 2001b/c), most frequent (from corpus) seen as most likely
	▪ Simplicity principle (Collard et al, 1981; LJ1983)	▪ Simplicity principle (Frazier, 1978, Bod, 2000) ▪ Involves perception and integration of various elements including melody, harmony, pitch, rhythm and timbre (Schmithorst, 2005) ▪ Prefer strong metrical position, harmonically consonant, the analysis that makes the best melodic and harmonic connections with adjacent time-spans, in choosing the head of a time-span, strongly prefer a cadential event that articulates the end of a group (JL1980)
	▪ Simplicity then Likelihood is most accurate combination (Bod, 2002) ▪ Syntactic Prediction Locality Theory (Gibson, 1998) provides a metric for structural integration and maintenance (mem) cost during syntactic processing	

Characteristics of stimuli (continued)		▪ A self-referential system that ignores the signifier-signified contrast (Besson & Friederici, 2005)
	▪ Grammaticality: is this string a sentence in the language in question? (JL1980)	▪ Grammaticality less important (JL1980)
	▪ Ambiguity: Can a string be assigned two or more structures with different meanings? (JL1980)	▪ Ambiguity is much more frequent, can often construe a passage in a multiplicity of ways (JL1980)
	▪ Tied down to specific meanings and functions (JL1980)	▪ Not tied down to specific meanings and functions (JL1980)
	▪ Metrical structure established in terms of the tree, tree chosen to optimise the resulting metrical pattern; flexible (JL1980)	▪ Metrical structure fixed in advance, time-span chosen so as to conform to it; fixed (JL1980)
	▪ Physically different but some of the computations involved in perceiving the meaning may share some similarities (Besson, 1999) ▪ Rule-based, require specific knowledge, indigenous to humans, sequential events that extend in time (Besson, 1999) ▪ Auditory, highly planned, internally consistent (Saffran, 2003) ▪ Perceptually discrete elements organised in hierarchically structured sequences (Jentschke, Koelsch, & Friederici, 2005) ▪ Both employ rhythmic and melodic patterns and are rule-governed (Patel, 2005) ▪ Conveyed by sounds, ubiquitous, specific to humans, cultural artefacts, rule-based systems composed of basic elements that are combined into higher-order structures through the rules of harmony and syntax (Besson & Schon, 2001) ▪ Hierarchical structuring (Tettamanti & Weniger, 2006) ▪ Perceptually discrete elements organised into hierarchically structured sequences (Patel, 2003), convey organised patterns of meaning ▪ Organised temporarily with the relevant structures unfolding in time (McMullen & Saffran, 2004) ▪ Reach our perceptual systems as frequency spectra (McMullen & Saffran, 2004) ▪ Some systematicities are universal, others are culture-specific (McMullen & Saffran, 2004) ▪ Sounds organised into discrete categories facilitating representation and memory (McMullen & Saffran, 2004) ▪ Infinitely combinatorial (McMullen & Saffran, 2004) ▪ Strong divergences at level of meaning, but significant parallels with respect to grammar and intonation (Brown et al., 2006) ▪ Generative phrasing systems in which phrases have larger meanings than the sum of their elements (Brown et al., 2006) ▪ Combinatorial, large structures generated hierarchically from a pool of smaller more unitary components (Brown et al., 2006) ▪ Can convey emotions through variation in pitch, amplitude, tempo, rhythm (Brown et al., 2006) ▪ Poetry in language and music both make use of basic prosodic cues but also requires cultural and syntactic knowledge for full appreciation (McMullen & Saffran, 2004) ▪ Discrete elements ordered into hierarchical patterns according to certain principles of combination (Raffman, 1993) ▪ Limited pool of discrete building blocks arranged combinatorially to generate structured phrases modulated by expressive phrasing mechanisms (Brown, 2001) ▪ Shared properties (vocalisation, affective prosody) Parallel (discreteness, combinatoriality, phrase formation, phrasing) Distinct (isometric rhythms, pitch blends, words, propositional syntax) (Brown, 2001) ▪ Some structures are present cross-culturally, some vary across cultures (Saffran, 2003)	

Processing	▪ Different modules (syntax, semantics, pragmatics) independent from each other? (Besson, 1999) ▪ Evidence for use of memory in probability models (Saffran, 2000) ▪ Broca involved in synt processing (Brown et al., 2006)	▪ While some networks overlap with ones used for language processing, music processing may involve its own domain-specific processing subsystems (Schmithorst, 2005) ▪ Evidence for use of memory in probability models (Jurafsky, 2002) ▪ Broca involved suggesting that sequential ordering of music may be processed in a homologous manner to language syntax (Brown et al., 2006) ▪ Music understood on the basis of foreground-background relationships, conveying meaning, created between the actual music and the underlying meter (Vuust, Roepstorff, Wallentin, Mouridsen, & Ostergaard, 2006) ▪ The perceptual parser hears each piece of music for the "first time" each time it is encountered (Jackendoff, 1991) ▪ A target chord is processed more rapidly if it is close to the tonal centre created by the prime (Patel, 2005) in musicians and non-musicians in western culture, showing implicit knowledge
	▪ Linguistic and musical syntactic processing engage different cognitive operations, but rely on a common set of neural resources for processes of structural integration in working memory (Patel, 1998) ▪ Shared Syntactic Integration Resource Hypothesis (Patel, 2003) ▪ Syntax in language and music share a common set of processes that operate on different structural representations (Patel, 2003) ▪ Commonalities at level of input (basic auditory processing) but significant differences with regard to domain-specific processing (Brown et al., 2006) ▪ Concatenating simpler memorised acoustic units (Brown et al., 2006) ▪ Information processing relies on keeping track of structural dependencies and performing structural integration (Tettamanti & Weniger, 2006) ▪ Experienced listeners can show implicit knowledge of syntactic patterns and principles: judgements of correctness, memory advantage for rule-governed sequences (Patel, 1998) ▪ Structural ambiguity and resolution in perception of unfolding sequence (Patel, 1998) ▪ Context-dependent psychological functions that are purely relational (subject / predicate; tonic / dominant) (Patel, 1998) ▪ Perception crucially depends on mem and integration in perception of structural relations between elements (Patel, 1998), the mind must integrate incoming elements with remembered previous element sin order to understand their place in the unfolding network of structural relations ▪ Lower levels of perceptual analysis become automated (and hence unconscious) in music and language (discontinuous dependencies) (Swinney & Love, 1998) ▪ Statistical learning works in both (Saffran, 2003) ▪ Internalise regularities by passive exposure; learned in incidental manner, used elaborately by adults (Jentschke et al., 2005) ▪ The syntactic structure not only provides the parsing systemto build up structural hierarchies and relationships among various phrases, but it also allows more information to be kept in memory (Jentschke et al., 2005)	

Automatic correcting	▪ Processing of dysfluencies	▪ Substitute correct reading based on knowledge of the context (Longuethiggins, 1994)
	▪ Production of plausible substitutes during imperfect recall (Patel, 1998)	
Sources of ambiguity / difficulty / violations	▪ Difficulty associated with long-distance integrations or integrations that force a revision of the current syntactic structure (Patel, 1998)	▪ Rhythm grouping (Longuethiggins, 1994) ▪ Tonal context (Longuethiggins, 1994) ▪ Difficult integrations are likely to be those that are unexpected because of melodic, harmonic, or rhythmic factors (Patel, 1998)
Cognitive effects of ambiguity		▪ "by supplying a different tonal context, the chord has been made to flip rather like the Necker cube" (Longuethiggins, 1994)
Expectancies and probabilities	▪ DLT accounted for within activation-based framework (reactivate dependent word which enables interpretation of context for input) (Patel, 2003)	▪ Parameter estimation from corpus (Ponsford et al., 1999) ▪ TPS accounted for in activation-based framework: the greater the tonal distance between the incoming and context chords, the greater the processing cost, presumably because the incoming chord was not predicted and thus had a low activation level (Patel, 2003) ▪ Musical expectation linked to tonal hierarchies (See Beinsteiner et al, 1999 and Schinuckler, 1989) ▪ Rhythm establishes a platform for expectancies (Vuust et al., 2006) ▪ The organisation of sound is relational, creating expectations; violations of musical expectancy are tantamount to violations of musical syntax (Lim, 2006) ▪ Musicians generally have different expectations of what might precede and follow a particular note depending on its pitch name (McLane, 1996); rhythm notation has same effect ▪ Musical training increases ERAN suggesting that more specific representations of musical regularities lead to heightened musical expectancies (Jentschke et al., 2005) ▪ Musical events point to and make us expect another musical event (Meyer) ▪ Musical expectancies automatically generated during the perception of a musical context (REFS 3-9 in (Jentschke et al., 2005))

Expectancies and probabilities (continued)	▪ "strong expectancies develop while listening to speech and music: as a specific word is expected within a specific linguistic context, a specific note or chord is expected within a musical context" (Besson, 1999) ▪ Both systems generate strong expectancies (Besson & Schon, 2001) just as a specific oword is expected within a specific linguistic context, specific notes or chords are expected at a given moment within a musical phrase ▪ Syntactic Prediction Locality Theory states that syntactic parsing involves simultaneous processes of prediction and structural integration. Prediction is elicited by knowledge of what word categories are required to complete a grammatical sentence given the input up to the current moment; maintaining these predictions until the words are encountered is associated with a memory cost (Patel, 1998) ▪ ERAN not seen in children with language impairment (Jentschke et al., 2005)	
Integration cost	▪ Distance and cost of integration (proportional to distance → numerical prediction) based on dependencies (Patel, 2003); Dependency Locality Theory	▪ Distance and cost of integration based on Lerdahl's tonal pitch space theory; compute the distance of each chord from the chord to which it attaches in the tree, with the added assumption that a chord inherits distances from the chords under which it is embedded (Patel, 2003)
	▪ DLT and TPS share the notion that structural integration is a key part of syntactic processing; that is mentally connecting each incoming element X to another element Y in the evolving structure (Patel, 2003); more distance, more difficult ▪ The greater the number of possible conformations that need to be discarded from the set of all possible conformations, the higher the computational cost; cost depends on degrees of freedom (Tettamanti & Weniger, 2006)	
ERP patterns and brain areas	▪ SEMANTICS: N400 inversely proportional to extent to which a word is expected (Kutas & Hillyard, 1980) ▪ Anterior negativities are language specific (Martin-Loeches, Casado, Gonzalo, De Heras, & Fernandez-Frias, 2006) ▪ SYNTAX: P600, index of analysis of formal, rule-governed aspects of language, proportional to degree of violation (Besson, 1999) ▪ Violation results in ELAN (Jentschke et al., 2005) ▪ Reflects the composite activity of multiple independent generators, each of which is responsible for a separate sub-process; LPCs reflecting overall structural integration cost (Martin-Loeches et al., 2006) ▪ Simple sentence < Garden path sentence < Ungrammatical sentence (Besson and Faita, 1995) ▪	▪ Corresponds to LPC/P300 in music, N400 is language specific (Besson and Macar, 1987) ▪ Corresponds to harmonic, same pattern in amplitude of LPCs (Patel et al, 1998) ▪ Violation results in ERAN (Jentschke et al., 2005) ▪ Equivalent to: Chord from same key < from near key < from distant key (Besson and Faita, 1995) ▪ Harmonic: deviant note at end of phrase LPC, ≃ P600 (Besson, 1999); expectancy determined through look at tonal hierarchy (LJ1983) ▪ Melodic: P300 in response to tritons; LPC/P300 for note within key but not correct one in familiar tune, of intermediate amplitude between out of key note and correct note (Besson, 1999) ▪ Rhythmic: long delay of final note caused large biphasic neg-pos complex elicited at time when

ERP patterns and brain areas (continued)	▪ ▪ Temporal processing (Besson and Faita, 1995) slow neg component	▪ not should have been played, "emitted potential" (Sutton at al, 1977), larger for familiar than unfamiliar; N1P2 elicited when note is finally played (Besson, 1999) ▪ Temporal processing (Besson and Faita, 1995) large emitted potential (but difference due to modality rather than to domain?
	▪ BA 47 involved in general neuronal processing of temporal processing subserving both language and music (Vuust et al., 2006) ▪ Activity in BA 40 involved in coping with stimuli that allow for more than one interpretation (Vuust et al., 2006)	
Psychological/emotional effects of ambiguity / violation of current understanding		▪ "a truly exhilarating experience" (Longuethiggins, 1994) ▪ Expectancies through rhythm leading to tension and relief created in different layers of the musical structure (background vs foreground) (Jones et al, 2002) ▪ Experience of metric tension as one of the central meaningful structure in music (Vuust et al., 2006) ▪ Looked at how modulations in the anticipatory structure allow for the perception of a series of auditory events as meaningful music (Vuust et al., 2006) ▪ Rules violated for artistic and aesthetic purposes (Lim, 2006)

Appendix B

Stimulus set

B.1 Stimulus lists, pre- and post-selection

B.1.1 *Harmony* stimuli

Stimulus set	Song	Subtype	Key	Tempo	Total Score	Selected	New Key
h1a1	Helly goodbye	a	B	100	7	✓	B
h1a2	Here comes the sun	a	A	70	-1	×	
h1a3	Something	a	Eb	130	4	✓	Eb
h1b1	Day in the life	b	Ab	70	-3	×	
h1b2	Dig a pony	b	F#	130	11	✓	F#
h1b3	With a little help	b	C	100	-2	×	
h1c1	Lady madonna	c	G	70	9	✓	G
h1c2	Magical mystery tour	c	F	130	12	✓	F
h1c3	Penny lane	c	D	100	5	✓	D
h2a1	Girl	a	C	100	5	✓	C
h2a2	Nowhere man	a	D	130	11	✓	D
h2a3	Places I remember	a	Bb	70	4	✓	Bb
h2b1	All you need is love	b	G	130	6	✓	G
h2b2	Help	b	C#	70	-3	×	
h2b3	Hey Jude	b	F	100	2	✓	F
h2c1	Here comes the sun	c	A	130	3	✓	A
h2c2	Penny lane	c	Ab	100	-3	×	
h2c3	All my loving	c	E	70	6	✓	E
h3a1	I wanna hold your hand	a	G	130	12	✓	Ab *
h3a2	Something	a	Ab	70	-1	×	
h3a3	Yesterday	a	C#	100	3	✓	C#
h3b1	Hey Jude	b	Eb	130	0	×	
h3b2	Norwegian wood	b	E	100	1	✓	E
h3b3	Places I remember	b	B	70	-2	×	
h3c1	Dear Prudence	c	C	100	0	×	
h3c2	Something	c	F#	130	9	✓	F#
h3c3	Try to see it my way	c	Bb	70	2	✓	Bb
h4a1	I will	a	F	100	-1	×	
h4a2	Mother nature's son	a	E	70	4	✓	C# *
h4a3	Something	a	F#	130	10	✓	Ab *
h4b1	Dear Prudence	b	D	130	2	✓	C *
h4b2	Piggies	b	Bb	100	8	✓	A *
h4b3	Yesterday	b	A	70	-3	×	
h4c1	All you need is love	c	Eb	130	9	✓	Eb
h4c2	Dear Prudence	c	C#	70	-3	×	
h4c3	Hard day's night	c	B	100	4	✓	B

B.1.2 *Rhythm* stimuli

Stimulus set	Song	Subtype	Key	Tempo	Total Score	Selected	New Key
r1a1	If there's anything	a	A	100	-2	×	
r1a2	Something	a	B	130	1	✓	B
r1a3	With a little help	a	Eb	70	3	✓	Eb
r1b1	I me mine	b	Ab	70	7	✓	Ab
r1b2	I've got a feeling	b	C	100	2	✓	C
r1b3	She loves you	b	F#	130	3	✓	F#
r1c1	Across the universe	c	C#	130	1	✓	C *
r1c2	Hey Jude	c	Bb	100	5	✓	Bb
r1c3	Long long long	c	F	70	-4	×	
r2a1	8 days a week	a	Bb	70	5	✓	Bb
r2a2	Can't buy me love	a	C	130	1	×	
r2a3	Michelle	a	D	100	4	✓	D
r2b1	Girl	b	C#	70	9	✓	C#
r2b2	Ticket to ride	b	F	100	8	✓	F
r2b3	Try to see it my way	b	G	130	3	✓	G
r2c1	Across the universe	c	A	130	-3	×	
r2c2	Lucy	c	Ab	70	5	✓	Ab
r2c3	Revolution	c	E	100	1	×	
r3a1	Get back	a	Ab	70	4	✓	A *
r3a2	Glass onion	a	C#	100	8	✓	C#
r3a3	Here comes the sun	a	G	130	3	✓	G
r3b1	Help	b	B	130	0	×	
r3b2	I me mine	b	E	100	7	✓	E
r3b3	Love me do	b	Eb	70	7	✓	Eb
r3c1	Across the universe	c	C	130	0	×	
r3c2	Glass onion	c	D	70	-5	×	
r3c3	Happiness	c	F#	100	4	✓	F#
r4a1	Don't let me down	a	E	130	2	✓	E
r4a2	Happiness	a	F	100	4	✓	F
r4a3	Martha my dear	a	F#	70	-2	×	
r4b1	Dear prudence	b	A	70	4	✓	A
r4b2	Hey Jude	b	Bb	130	11	✓	B *
r4b3	Long long long	b	D	100	-3	×	
r4c1	Fool on a hill	c	B	100	-5	×	
r4c2	Glass onion	c	Eb	70	1	✓	D *
r4c3	Ticket to ride	c	G	130	-1	×	

B.1.3 *Semantics* stimuli

Stimulus set	Subtype	Total Score	Selected
sea1	a	12	✓
sea2	a	9	×
sea3	a	13	✓
sea4	a	10	✓
sea5	a	7	×
sea6	a	9	×
sea7	a	13	✓
sea8	a	11	✓
sea9	a	12	✓
sea10	a	2	×
sea11	a	12	✓
sea12	a	11	✓
seb1	b	9	×
seb2	b	9	✓
seb3	b	2	×
seb4	b	12	✓
seb5	b	12	✓
seb6	b	5	×
seb7	b	11	✓
seb8	b	14	✓
seb9	b	9	×
seb10	b	10	✓
seb11	b	7	×
seb12	b	9	✓
sec1	c	11	✓
sec2	c	10	✓
sec3	c	3	×
sec4	c	8	×
sec5	c	11	✓
sec6	c	13	✓
sec7	c	11	✓
sec8	c	12	✓
sec9	c	10	✓
sec10	c	10	✓
sec11	c	9	×
sec12	c	10	✓

B.1.4 *Syntax* stimuli

Stimulus set	Subtype	Total Score	Selected
sea1	a	12	✓
sea2	a	9	×
sea3	a	13	✓
sea4	a	10	✓
sea5	a	7	×
sea6	a	9	×
sea7	a	13	✓
sea8	a	11	✓
sea9	a	12	✓
sea10	a	2	×
sea11	a	12	✓
sea12	a	11	✓
seb1	b	9	×
seb2	b	9	✓
seb3	b	2	×
seb4	b	12	✓
seb5	b	12	✓
seb6	b	5	×
seb7	b	11	✓
seb8	b	14	✓
seb9	b	9	×
seb10	b	10	✓
seb11	b	7	×
seb12	b	9	✓
sec1	c	11	✓
sec2	c	10	✓
sec3	c	3	×
sec4	c	8	×
sec5	c	11	✓
sec6	c	13	✓
sec7	c	11	✓
sec8	c	12	✓
sec9	c	10	✓
sec10	c	10	✓
sec11	c	9	×
sec12	c	10	✓

B.1.5 Transcribed *semantics* stimuli

Example stimulus triplet

Incongruous-Resolved Lizzie's friendship group was an oasis in which she found comfort from the trials of everyday life.

Normal Ein Gedi is a famous oasis in the desert to the west of the Dead Sea.

Incongruous-Unresolved Lizzie's friendship group was an oasis in the desert to the west of the Dead Sea.

B.1.5.1 Simple metaphor

1. **Wall**

   Incongruous-Resolved In tennis matches Chris was a wall which could bounce all the balls back over the net.

   Normal In each classroom there was a wall which could be used to display all the pupils' pictures.

   Incongruous-Unresolved In tennis matches Chris was a wall which could be used to display all the pupils' pictures.

2. **Feather**

   Incongruous-Resolved The neighbour's dog was a feather that even small children could pick up.

   Normal What the teacher brought in was a feather that even the youngest children recognised as a peacock feather.

   Incongruous-Unresolved The neighbour's dog was a feather that all the children knew came from a peacock.

3. **Fortress**

Incongruous-Resolved The supportive family was a fortress that Henry knew he could always rely on.

Normal The Tower of London was a fortress that Henry VIII used as a prison in the 16th century.

Incongruous-Unresolved The supportive family was a fortress that Henry VIII used as a prison in the 16th century.

4. **Motorway**

Incongruous-Resolved The ant trail was the motorway that the insects used to transport food.

Normal The M4 was the motorway that the family took when they visited Mrs Jones.

Incongruous-Unresolved The ant trail was the motorway that the a the family took when they visited Mrs Jones.

5. **Encyclopedia**

Incongruous-Resolved Because of his passion for flowers Nathanael was an encyclopedia that everyone loved going hiking with.

Normal The latest addition to the company's website was an encyclopedia that everyone could edit online.

Incongruous-Unresolved Because of his passion for flowers Nathanael was an encyclopedia that everyone could edit online.

6. **Maze**

Incongruous-Resolved The psychology department is a maze that the new students struggle to find their way around.

Normal The centrepiece of the Chateau's garden is a maze that the gardeners have to trim regularly.

Incongruous-Unresolved The psychology department is a maze that the gardeners have to trim regularly.

7. **Shackles**

Incongruous-Resolved Rachel's supervisors were concerned that her perfectionism might turn into shackles which would stop her from finishing her research on time.

Normal This is the place in the old prison where the convicts were put into shackles which would stop them from escaping even if the doors were left unlocked.

Incongruous-Unresolved Rachel's supervisors were concerned that her perfectionism might turn into shackles which would stop them from escaping even if the doors were left unlocked.

8. **Cage**

Incongruous-Resolved Lucy's depression soon turned into a cage that prevented her from thinking straight.

Normal The rats were placed into a cage that prevented them from running around the lab.

Incongruous-Unresolved Lucy's depression soon turned into a cage that prevented them from running around the lab.

9. **Missile**

Incongruous-Resolved The young kitten was a missile that could chase after anything that moved.

Normal The army's favorite weapon was a missile that could be launched across the Atlantic.

Incongruous-Unresolved The young kitten was a missile that could be launched across the Atlantic.

10. **Abyss**

Incongruous-Resolved Paul's untidy room was an abyss into which none of his housemates dared venture.

Normal In the middle of the landscape there was an abyss into which some streams had formed waterfalls.

Incongruous-Unresolved Paul's untidy room was an abyss into which some streams had formed waterfalls.

11. **Crutch**

Incongruous-Resolved In the wake of the family crisis, James was a crutch that Susan could rely on to comfort her in emotional times.

Normal The main thing the family still needed from the hospital was a crutch that Susan could use to get up the stairs.

Incongruous-Unresolved In the wake of the family crisis, James was a crutch that Susan could use to get up the stairs.

12. **Armour**

Incongruous-Resolved When Jacob's friends left him, his belief was the armour with which he withstood life's toughest trials.

Normal The most famous piece in the museum was the armour with which Henry V fought in the battle of Agincourt.

Incongruous-Unresolved When Jacob's friends left him, his belief was the armour with which Henry V fought in the battle of Agincourt.

B.1.5.2 Metaphorical adjectives

1. **Tepid**

   Incongruous-Resolved Because his speech was so tepid, Joseph struggled to recruit more people for the task.

   Normal Because his meal was so tepid, Joseph asked the staff to reheat the vegetables.

   Incongruous-Unresolved Because his speech was so tepid, Joseph asked the staff to reheat the vegetables.

2. **Spicy**

   Incongruous-Resolved Luke's outfit was so spicy that no one else was noticed by the talent scout.

   Normal The main course was so spicy that no one could taste the desert.

   Incongruous-Unresolved Luke's outfit was so spicy that no one could taste the desert.

3. **Aching**

   Incongruous-Resolved When the members of the congregation heard the tragic announcement their hearts began to ache as they thought of all the friends and families of the victims.

   Normal Despite all the training the athletes had undertaken, their muscles began to ache as they entered into the last kilometer of the race.

   Incongruous-Unresolved When the members of the congregation heard the tragic announcement their hearts began to ache as they as they entered into the last kilometer of the race.

4. **Bland**

Incongruous-Resolved All the delegates agreed that Harry's speech was very bland because he had not been able to find a powerful illustration for his main point.

Normal All the guests complained that Harry's Mexican dish was very bland because he had forgotten to add the chilli powder to the sauce.

Incongruous-Unresolved All the delegates agreed that Harry's speech was very bland because he had had forgotten to add the chilli powder to the sauce.

5. **Deafening**

Incongruous-Resolved I remember John saying that after all the noise of the party the silence in the deserted streets had been deafening as he walked home alone.

Normal The sound engineer apologised for the noise that had been deafening as he should have checked the level of the speakers properly.

Incongruous-Unresolved I remember John saying that after all the noise of the party the silence in the deserted streets had been deafening as he should have checked the level of the speakers properly.

6. **Brightest**

Incongruous-Resolved It is very sweet to hear Tom say that his relationship with Sarah is the brightest thing that he has in his life at the moment.

Normal Even when it is masked by clouds the sun is the brightest thing that plants can rely on for photosynthesis.

Incongruous-Unresolved It is very sweet to hear Tom say that his relationship with Sarah is the brightest thing that plants can rely on for photosynthesis.

7. **Quiet**

Incongruous-Resolved Despite all of Mark's ongoing worries, his mind was completely quiet last night and he managed to go to sleep as soon as he went to bed.

Normal Although there usually is a lot of traffic on Jake's street, it was completely quiet last night because the lorries had been diverted onto a parallel street.

Incongruous-Unresolved Despite all of Mark's ongoing worries, his mind was completely quiet last night because the lorries had been diverted onto a parallel street.

8. **Rusty**

Incongruous-Resolved The old pianist said that his left hand had become rusty over time because he had not put enough effort into maintaining his virtuoso skills.

Normal The old cyclist had to admit that his bike had become rusty over time because he had left it out in the rain far too often.

Incongruous-Unresolved The old pianist said that his left hand had become rusty over time because he had left it out in the rain far too often.

9. **Engraved**

Incongruous-Resolved Despite the old man's many years, his mind was still engraved with the beautiful memories of his youth.

Normal Despite its old age, the silver bracelet was still engraved with the initials of the jeweller who had made it.

Incongruous-Unresolved Despite the old man's many years, his mind was still engraved with the initials of the jeweller who had made it.

10. **Torn**

Incongruous-Resolved The dark and gloomy night had been torn by a bolt of lightening.

Normal The front of the library book had been torn by a pupil last year.

Incongruous-Unresolved The dark and quiet night had been torn by a pupil last year.

11. **Foggy**

Incongruous-Resolved After the incident Tim said that his memory was slightly foggy because he couldn't quite remember what had happened.

Normal That morning the skier knew that the day was going to be slightly foggy because he could't see the highest peak in the distance.

Incongruous-Unresolved After the incident she said that her memory was slightly foggy because he could't see the highest peak in the distance.

12. **Spiky**

Incongruous-Resolved The new boss' personality was quite spiky and the employees were all afraid of his nasty comments.

Normal The Christmas tree we bought last year was quite spiky and the decorations didn't stay on it very easily.

Incongruous-Unresolved The new boss' personality was quite spiky and the decorations didn't stay on it very easily.

B.1.5.3 Submerged or implied metaphor

1. **Sprouting**

Incongruous-Resolved After the phone call, a huge smile was seen sprouting from the corners of Claire's mouth.

Normal After the rain, a new weed was seen sprouting from the corners of the flower bed.

Incongruous-Unresolved After the phone call, a huge smile was seen sprouting from the corners of the flower bed.

2. **Clanking**

Incongruous-Resolved William was sure that his mind could be heard clanking as he tried desperately to think of something to say.

Normal William knew it would not be long before his car could be heard clanking as he had been warned that he needed a new exhaust pipe.

Incongruous-Unresolved William was sure that his mind could be heard clanking as he had been warned that he needed a new exhaust pipe.

3. **Scream**

Incongruous-Resolved Though Laura could not speak, her eyes were constantly screaming for the love and attention she so dearly longed for.

Normal The teacher was growing tired of the children who were constantly screaming for the sake of disrupting the lessons.

Incongruous-Unresolved Though Laura could not speak, her eyes were constantly screaming for the sake of disrupting the lessons.

4. **Chime**

Incongruous-Resolved As Christmas time approached, the smell of pine trees had began to chime in the little girl's lungs.

Normal As the wedding ceremony came to an end, the church bell had began to chime in the little village square.

Incongruous-Unresolved As Christmas time approached, the smell of pine trees had began to chime in the little village square.

5. **Incinerate**

Incongruous-Resolved Andrew's simple words of encouragement rapidly incinerated all her deepest fears.

Normal Because the fire in the house was so fierce, it rapidly incinerated all her possessions.

Incongruous-Unresolved Andrew's simple words of encouragement rapidly incinerated all her possessions.

6. **Eating**

Incongruous-Resolved Harold's bitterness was slowly eating up all the fond memories he had of Belinda.

Normal The huge slugs were slowly eating up all the plants in Belinda's garden.

Incongruous-Unresolved Harold's bitterness was slowly eating up all the fond memories he had of Belinda.

7. **Suffocate**

Incongruous-Resolved The old firm was gradually suffocating because of a lack of new intakes.

Normal The prisoners were gradually suffocating because of a lack of air.

Incongruous-Unresolved The old firm was gradually suffocating because of a lack of air.

8. **Bud**

Incongruous-Resolved Their relationship was starting to bud again after a very difficult few months.

Normal The plants in the garden were starting to bud again thanks to all the rain this summer.

Incongruous-Unresolved Their relationship was starting to bud again thanks to all the rain this summer.

9. **Blossom**

   Incongruous-Resolved Helen's family had not expected her to have blossomed so much at university because of the tough time she had been through in her first term.

   Normal The Gavins had not expected the cherry tree to have blossomed so much because of the frost that had happened earlier this year.

   Incongruous-Unresolved Helen's family had not expected her to have blossomed so much because of the frost that had happened earlier this year.

10. **Bite**

   Incongruous-Resolved Although he had meant no harm, James' comment to Sally had bitten her quite hard and she was struggling to forgive him.

   Normal The vets all agreed that the dog that had attacked Sally had bitten her quite hard and that it had to be put down.

   Incongruous-Unresolved Although he had meant no harm, James' comment to Sally had bitten her quite hard and that it had to be put down.

11. **Climb**

   Incongruous-Resolved After what had seemed like a very long night, daylight could be seen climing the walls of the houses as the sun rose over the city.

   Normal After the introduction of healthy school meals, children could be seen climbing the walls of some schools to escape and go to the chippy.

   Incongruous-Unresolved After what had seemed like a very long night, daylight could be seen climing the walls of some schools to escape and go to the chippy.

12. **Accompany**

Incongruous-Resolved The boss' son was so full of himself that only his pride would ever accompany him to the end of year party.

Normal The young man wondered whether his girlfriend would ever accompany him to London to meet his parents.

Incongruous-Unresolved The boss' son was so full of himself that only his pride would ever accompany him to London to meet his parents.

B.1.6 Transcribed *syntax* stimuli

Example stimulus triplet

Incongruous-Resolved: It is unusual to see businessmen motorbike their way to the office.

Normal: Although Lucy's car was never clean, her motorbike was always spotless.

Incongruous-Unresolved: It is unusual to see businessmen motorbike was always spotless.

B.1.6.1 Locatum verbs

1. **Cinnamon**

   Incongruous-Resolved To add a bit of taste the Smiths cinnamon their hot chocolate whenever they get a chance to.

   Normal Because there was something special about the Smiths' cinnamon their hot chocolate with spices always tasted amazing.

   Incongruous-Unresolved To add a bit of taste the Smiths cinnamon their hot chocolate with spices always tasted amazing.

2. **Glove**

   Incongruous-Resolved In keeping with modern health and safety rules, ice-hockey players glove their hands to prevent nasty cuts.

Normal Although the team members all dreamed of wearing the legendary baseball player's glove their hands were the wrong size for it.

Incongruous-Unresolved In keeping with modern health and safety rules, ice-hockey players glove their hands were the wrong size for it.

3. **Blanket**

Incongruous-Resolved Because it gets very cold in old farm houses the French blanket their bed in winter.

Normal Although the Duponts' room was very cold under the French blanket their bed was warm.

Incongruous-Unresolved Because it gets very cold in old farm houses the French blanket their bed was warm.

4. **Sheet**

Incongruous-Resolved To prevent any damage the painters sheet the furniture in the room.

Normal The owner knew that under the painter's sheet the furniture would be safe.

Incongruous-Unresolved To prevent any damage the painters sheet the furniture would be safe.

5. **Trouser**

Incongruous-Resolved Because her two-year-old never stops kicking Mrs Matthews trousers her son with great difficulty every morning.

Normal Because Mary had accidentally ironed a hole in Matthew's trousers her son had to wear shorts to school.

Incongruous-Unresolved Because her two-year-old never stops kicking Mrs Matthews trousers her son had to wear shorts to school.

6. **Cloak**

    Incongruous-Resolved To keep the new collection a secret the designer cloaks the models all the way to the catwalk.

    Normal All the fashion critics said that in the designer cloaks the models looked stunning.

    Incongruous-Unresolved To keep the new collection a secret the designer cloaks the models looked stunning.

7. **Uniform**

    Incongruous-Resolved The manager of IKEA uniforms the staff in yellow so that they can easily be spotted by customers.

    Normal In the new IKEA uniforms the staff look very professional but nobody is sure why they are dressed in yellow.

    Incongruous-Unresolved The manager of IKEA uniforms the staff look very professional but nobody is sure why they are dressed in yellow.

8. **Tomato**

    Incongruous-Resolved Thankfully we have not yet reached as day where pupils tomato their teachers as a sign of protest.

    Normal Although the biology class had put a lot of effort into the growth of the pupils' tomato, their teacher's plant was growing much faster than theirs.

    Incongruous-Unresolved Thankfully we have not yet reached a day where pupils tomato their teachers plant was growing much faster than theirs.

9. **Lemon**

    Incongruous-Resolved Instead of adding milk the lebanese lemon their tea to add some taste.

Normal Because all the hot drinks in the restaurant came with a dash of lebanese lemon their tea tasted exceptionally good.

Incongruous-Unresolved Instead of adding milk the lebanese lemon their tea tasted exceptionally good.

10. **Cheese**

Incongruous-Resolved The owner of the restaurant objects to the fact that the new chefs cheese all the dishes on the menu.

Normal Although nobody really liked the new chef's cheese all the customers liked his homemade yoghurts.

Incongruous-Unresolved The owner of the restaurant objects to the fact that the new chefs cheese all the customers liked his homemade yoghurts.

11. **Statue**

Incongruous-Resolved Generally speaking, gardeners who work for emperors statue every landscape they create.

Normal When the courtiers were shown the emperor's statue every one of them bowed down in awe to the artist.

Incongruous-Unresolved Generally speaking, gardeners who work for emperors statue every one of them bowed down in awe to the artist.

12. **Gnome**

Incongruous-Resolved To add a bit of fun to their property the neighbours gnome their garden in the summer.

Normal If someone had the good taste to remove neighbours' gnome their garden would look amazing.

Incongruous-Unresolved To add a bit of fun to their property the neighbours gnome their garden would look amazing.

B.1.6.2 Location verbs

1. **Sofa**

   Incongruous-Resolved After a long day at work, the Thompsons sofa themselves and play a game of scrabble.

   Normal Although they had built the Thompsons' sofa themselves the staff of the furniture firm were incapable of fixing it.

   Incongruous-Unresolved After a long day at work, the Thompsons sofa themselves the staff of the furniture firm were incapable of fixing it.

2. **Barn**

   Incongruous-Resolved In heavy rain the farmers barn the cows to keep them warm.

   Normal The people knew that in this farmer's barn the cows were well looked after.

   Incongruous-Unresolved In heavy rain the farmers barn the cows were well looked after.

3. **Footnote**

   Incongruous-Resolved When there is no room in the main body of the text authors footnote lengthy explanations to ease comprehension.

   Normal Although the students all claimed to have read the author's footnote lengthy discussions revealed that they hadn't understood the paper.

   Incongruous-Unresolved When there is no room in the main body of the text authors footnote lengthy discussions revealed that they had not understood the paper.

4. **Porch**

Incongruous-Resolved In all good American films paperboys porch the newspapers as they cycle past.

Normal The news-deprived neighbourhood spotted on the paperboy's porch the newspapers he had forgotten to deliver.

Incongruous-Unresolved In all good American films paperboys porch the newspapers he had forgotten to deliver.

5. **Attic**

Incongruous-Resolved When there is no room in conventional offices some departments attic the youngest members of staff.

Normal Although there is not much room in the department's attic the youngest members of staff enjoy working there.

Incongruous-Unresolved When there is no room in conventional offices the department's attic the youngest members of staff enjoy working there.

6. **Safe**

Incongruous-Resolved Whenever they go on holiday the Taylors safe the expensive jewelery to be sure it will not go missing.

Normal The suit-making apprentice was told that if he put his creations in the tailor's safe the expensive fabric would not be damaged.

Incongruous-Unresolved Whenever they go on holiday the tailor's safe the expensive fabric would not be damaged.

7. **Plate**

Incongruous-Resolved To improve the speed with which customers are served in this restaurant, the waitresses plate the food themselves so that the chefs can concentrate on the cooking.

Normal Though the restaurant fed all the staff, it was true that on a waitresses' plate the food was never as nicely presented as on a customers' plate.

Incongruous-Unresolved To improve the speed with which customers are served in this restaurant, the waitresses plate the food was never as nicely presented as on a customers' plate.

8. **Doormat**

Incongruous-Resolved Usually, proud house owners doormat their shoes on the way in, to prevent any mud from entering the house.

Normal Although the tennants had all used the owner's doormat their shoes had left some mud on the white carpet.

Incongruous-Unresolved Usually, proud house owners doormat their shoes had left some mud on the white carpet.

9. **Mantlepiece**

Incongruous-Resolved To honour their children's academic achievement some parents mantlepiece really odd-looking graduation photos.

Normal The children said that their parents' mantlepiece really didn't fit with the rest of the decor in the front room.

Incongruous-Unresolved To honour their children's academic achievement some parents mantlepiece really didn't fit with the rest of the decor in the front room.

10. **Pinboard**

Incongruous-Resolved To make sure the information reaches all the members of staff the directors pinboard all the important notices.

Normal To make the relevant pieces of information stand out on the director's pinboard all the notices are colour-coded.

Incongruous-Unresolved To make sure the information reaches all the members of staff the directors pinboard all the notices are colour-coded.

11. **Jar**

Incongruous-Resolved It is often only when Andy jars his jam that he realises how little he has made.

Normal Although many people had given Andy jars his jam was always stored in plastic boxes.

Incongruous-Unresolved It is often only when Andy jars his jam was always stored in plastic boxes.

12. **Blackboard**

Incongruous-Resolved To help their pupils learn vocabulary, some language teachers blackboard important words.

Normal On my French teacher's blackboard important words were always written in green.

Incongruous-Unresolved To help their pupils learn vocabulary, some language teaches blackboard important words were written in green.

B.1.6.3 Instrument or agent/experiencer verbs

1. **Salt**

Incongruous-Resolved Karen knew that the only way to sort out the slug problem in the house was to salt the nasty creatures to death.

Normal Although he was not nearly as allergic to pepper as he was to salt the nasty reactions he got from salt put him off seasoning his food altogether.

Incongruous-Unresolved Karen knew that the only way to sort out the slug problem in the house was to salt the nasty reactions he got from salt put him off seasoning his food altogether.

2. **Pen-knife**

   Incongruous-Resolved Of all the fans of action movie heroes, only MacGuyver's penknife their way out of dangerious situations.

   Normal The characters knew that with the help of MacGuyver's penknife their way out of the situation would be straightforward.

   Incongruous-Unresolved Of all the action movie heroes' fans, only MacGuyver's penknife their way out of the situation would be straightforward.

3. **Towel**

   Incongruous-Resolved If he splashes too much water out of the bath Peter towels the wet floor to dry it before his mother notices.

   Normal Although the coach at the diving competition kept handing Peter towels the wet boy refused to dry himself.

   Incongruous-Unresolved When he's splashed too much water out of the bath Peter towels the wet boy refused to dry himself.

4. **Seatbelt**

   Incongruous-Resolved All responsible parents seatbelt small children to the car seats.

   Normal By playing with a parent's seatbelt small children have caused accidents.

   Incongruous-Unresolved All responsible parents seatbelt small children have caused accidents.

5. **Ruler**

   Incongruous-Resolved When she was at school, my grandmother often saw teachers ruler the disruptive pupil's hands as a punishment for their bad behaviour.

Normal Because they had stolen the teacher's ruler the disruptive pupils hoped the geometry lesson would be postponed.

Incongruous-Unresolved When she was at school, my grandmother often saw teachers ruler the disruptive pupil's hoped the geometry lesson would be postponed.

6. **Taxi**

Incongruous-Resolved Although it would not take long to walk there from the train station, some guests taxi to the hotel in order to save time.

Normal Although the wealthy couple had promised to refund the cost of their guest's taxi to the mansion this promise was not kept.

Incongruous-Unresolved Although it would not take long to walk there from the train station, some guests taxi to the mansion this promise was not kept.

7. **Lightsabre**

Incongruous-Resolved No one expected to see one of the fallen jedis lightsabre Obi-wan in such a callous way.

Normal Having previously managed to avoid the fallen jedi's lightsabre Obi-wan decided to give up the fight.

Incongruous-Unresolved No one expected to see one of the fallen jedis lightsabre Obi-wan decided to give up the fight.

8. **Bugle**

Incongruous-Resolved On one particular scout camp the leaders bugle the scouts out of their sleep at seven o'clock every morning.

Normal Having hidden the leader's bugle the scouts were hoping for a lie in this morning.

Incongruous-Unresolved On one particular scout camp the leaders bugle the scouts were hoping for a lie in this morning.

9. **Teaspoon**

Incongruous-Resolved Escape from Alcatraz is the famous film in which prisoners teaspoon their way to freedom.

Normal Although the guards had managed to retrieve the prisoner's teaspoon their way of obtaining it was frowned upon.

Incongruous-Unresolved Escape from Alcatraz is the famous film in which prisoners teaspoon their way of obtaining it was frowned upon.

10. **Hairpin**

Incongruous-Resolved The special agent that Bourne previously worked with hairpins any lock open in under 10 seconds.

Normal Ever since Ben had seen that some special agents who unlock doors worked with hairpins any lock his mother tried to open was stiff.

Incongruous-Unresolved The special agent that Bourne previously worked with hairpins any lock any lock his mother tried to open was stiff.

11. **Cheeseburger**

Incongruous-Resolved A recent report emphasised that parents should not let their teenagers cheeseburger themselves into obesity.

Normal Although the staff at the restaurant had not made the teenager's cheeseburger themselves they took full responsibility for the poor quality of the meat.

Incongruous-Unresolved A recent report emphasised that parents should not let their teenagers cheeseburger themselves they took responsibility for the poor quality of the meat.

12. **Blender**

    Incongruous-Resolved When their fruit starts getting old my flatmates blender the bananas into a nice smoothie.

    Normal With the help of my flatmate's blender the bananas were quickly turned into a nice smoothie.

    Incongruous-Unresolved When their fruit starts getting old my flatmates blender the bananas were quickly turned into a nice smoothie.

B.2 Sound files

Example sound files can be listened to at www.carafeatherstone.co.uk/research/stimuli. The entire stimulus set is available upon request from the author.