The Nature of Syntactic Processing in Music and Language

Anna Fiveash

BPsych(Hons), M.A.

Department of Psychology

ARC Centre of Excellence in Cognition and its Disorders

Macquarie University

Sydney, Australia

This thesis is submitted for the degree of PhD in Psychology

September, 2017

Table of Contents

Thesis Abstract

It has been suggested that music and language are processed with shared cognitive resources. As these processing resources are limited in capacity, the concurrent presentation of music and language should produce interference, such that reduced processing is observed in one or both domains. The aim of this thesis was to investigate shared *syntactic* processing in music and language. To this end, I conducted a series of experiments to address limitations in previous research on this topic, which has (1) depended on surprising violations of syntactic structure, which may have engaged shared non-syntactic processes between music and language; (2) ignored considerations of auditory streaming research; and (3) focused mainly on the effects of music syntax processing on language syntax processing but not vice versa. Chapter 1 outlines the theoretical basis for the thesis. Chapter 2 presents three experiments showing that syntactic interference can be observed without surprising violations of structure, and that syntactic processing is dependent on successful auditory streaming. Chapter 3 reports on an event-related potential (ERP) study suggesting that syntactic processing of music is reduced when auditory streaming is disrupted. Experiments in Chapters 4 and 5 suggest that syntactic interference from music to language is modulated by whether tasks are primary or secondary. In both chapters, syntactic interference was not observed on the primary tasks, but interference was observed on the secondary tasks. In Chapter 6, all the experimental findings are drawn together and interpreted within a new *Competitive Attention and Prioritisation Model.* This thesis provides a new understanding of the nature of syntactic processing in music and language, and provides insight into the simultaneous processing of syntax in these two important modes of human communication.

Statement of Candidate

I certify that the work presented in this thesis entitled: *The Nature of Syntactic Processing in Music and Language* has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree to any other university or institution other than Macquarie University. I also certify that this thesis is an original piece of research and it has been written by me. Any help or assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged. In addition I certify that all information sources and literature used are indicated in the thesis.

The research presented in this thesis was approved by the Macquarie University Human Ethics Committee (REF 5201500300). Some of the data presented in this thesis (Chapter 4, Experiment 1; Chapter 5, Experiment 1) was collected at the University of Toronto, Mississauga. Data collection was approved by the University of Toronto, Mississauga, Human Research Ethics Program (REF: NSERC20588), and was approved as offsite research at Macquarie University.

Signature:

Anna Fiveash (Student number: 43754260)

1st of September, 2017

Acknowledgements

I have had an incredible PhD journey, and it is largely due to the wonderful people that I have been surrounded with. I would first like to thank my principle supervisor, Bill Thompson. You have taught me so much throughout these three years, and I continue to be amazed by your insight and wisdom, and your ability to see the positives in every situation. You have been constantly supportive and encouraging, and you continue to inspire me to think bigger and better. To my associate supervisor - Genevieve McArthur - your support, encouragement, humour, and keen mind for logic have been invaluable throughout this process. I feel very lucky to have learned from you and your passion for science. I would also like to thank Nic Badcock for your time and patience teaching me how to code. Your passion for processes and coding is inspiring.

Thank you to the *Music, Sound, and Performance Lab* for weekly support, encouragement, and discussions. I would especially like to thank Xuejing Lu and Yanan Sun who were always willing to share their knowledge and skills. Thank you also to Olivia Brancatisano and Rebecca Gelding who have become close friends as well as colleagues. A big thank you to all the administrative staff who made my PhD run smoothly and offered me advice along the way, especially Lesley McKnight, Avril Moss, and Marcus Ockenden. Thank you also to Sachiko Kinoshita and Bianca DeWit for a great tutoring experience.

A huge thank you to my friends for making my time in Sydney filled with fun, support, adventures, dumplings, and wine. I am so grateful to have found such a wonderful group of people to call my friends. Thank you to Squad for all the adventures—Olivia Brancatisano, Rochelle Cox, Mariia Kaliuzhna, Kate Hardwick, and a special mention to Vana Webster, who has been extremely supportive throughout the whole process, and especially in the last few months. Thank you to Cassie Lyne for all the coastal walks and dumplings, and to my Canberra friends - Liz Flora, Hannah Studdert, Dion Pretorius, and Tabitha Hocking - thank you for the weekends away and for always being there. To everyone at Macquarie who has shared lunches, baked goods, and knowledge, thank you for creating such a wonderful environment to work in.

To my brother Rhys Fiveash, thank you for being awesome, I can't wait to see where you end up in life. And last but not least I am forever grateful to my amazing parents, Susanne Francisco and Mike Fiveash, for their continued love, support, encouragement, and enthusiasm throughout my PhD and whole life. I could not have done this without you.

Thesis by Publication

This thesis is written in the "thesis by publication" format. Chapters 2, 3, 4, and 5 have been written for independent publications in peer-reviewed journals. Chapters 2 and 3 are currently under review in separate peer-reviewed journals. In addition, tables and figures have been inserted within each chapter. Brief introductions at the beginnings of Chapters 2, 3, 4, and 5 provide explanations about how the particular chapter connects with other chapters in the thesis, and contributes to the broader themes of the thesis.

Chapter 1

Introduction

Chapter 1: Introduction

Chapter 1: Introduction

Music and language are two complex systems of communication that are ubiquitous in human cultures (Levinson, 2013; Mithen, 2009). Comparisons between music and language have a long history, with Darwin (1882) suggesting that language developed from musical sounds. From this early coupling of music and language, research has revealed commonalities and differences in the two domains. Modularity theories suggest that the brain processes music and language separately, and that they share no overlapping processing resources (e.g., Peretz & Coltheart, 2003; Piccirilli, Sciarma, & Luzzi, 2000). Shared processing theories posit that the brain uses the same processing resources for at least some features of music and language (e.g., Koelsch, 2013; Patel, 2008; Thompson, Marin, & Stewart, 2010). There is evidence supporting both modular and shared processing viewpoints, which has triggered much debate about the overlap and dissociation between music and language. Even within a shared processing approach, there are a number of fundamental questions and distinctions about which aspects of music and language do and do not share processing resources. This topic is further complicated by the roles that cognitive processes such as auditory streaming, working memory, and attention might play in processing music and language—none of which are well understood. Connections between music and language are therefore integral to the understanding of a number of important cognitive processes.

There are substantial and important differences between music and language that should be acknowledged in any discussion comparing the two domains. The most obvious distinction is that they are based on different elements that combine with each other to form larger meaningful units. Language consists of phonemes (a unit of speech that differentiates the meaning of one word from another, e.g., h*a*ppy versus h*i*ppy) that are combined into morphemes (the smallest meaningful unit of a language, e.g., happiness = happy + ness), and combined to form words (the smallest isolated unit of language with meaning, e.g., happiness). These words are then grouped to form sentences. In contrast to language, music consists of individual notes (discrete pitches) that are grouped together to form chords (a group of notes played simultaneously, e.g., the notes C, E, and G make up the C major chord), and notes and chords are grouped together to form musical phrases (e.g., complete sections of music such as a melody).

A second important difference is how language and music function to express meaning. Language can be used to communicate meaning that is usually unambiguous and *referential* (e.g., specific objects, events or ideas). For example, the sentence *let's go on a holiday* should communicate the same meaning to two separate listeners (Carnie, 2013; Jackendoff, 2009). Music, in contrast, acquires meaning in a way that is often ambiguous and flexible, such that listeners can interpret and use music in different ways depending on the social context. For example, Leonard Cohen's song "Hallelujah" was sung at the 2010 winter Olympics as a celebration of human achievement; the same song has been sung as a symbol of cultural inclusion, and in remembrance for victims of natural disasters. This quality has been labelled *floating intentionality* (Cross, 2014). Music can also be appreciated for the unfolding pattern of tension and relaxation that is generated by its internal logic and grammar—a property called *embodied meaning* (Meyer, 1956). In short, music and language consist of different elements, and the combination of these elements results in different communicative affordances.

Despite clear differences between music and language, there are also striking similarities between the two domains. Both music and language are primarily human attributes (Rebuschat, Rohrmeier, Hawkins, & Cross, 2012) that convey meaning to listeners (Scherer, 2013), are found in all cultures (Levinson, 2013), facilitate social bonding (Cross, 2014), and are suggested to share an evolutionary past (Brown, 2000; Mithen, 2007). In their auditory forms, they share a number of acoustic similarities including fluctuations in pitch, timbre, and intensity (Patel, 2008). In addition to acoustic similarities, music and language have syntactic structure on multiple hierarchical levels. *Syntax* refers to the set of rules or regularities that describe how discrete elements, such as notes and words, are combined into larger units such as phrases, and how those larger units are combined to form yet larger meaningful units such as melodies and sentences (Chomsky, 1957; Koelsch, 2013; Lerdahl & Jackendoff, 1983; Patel, 2008). Syntactic sequences are hierarchical (e.g., larger elements encompass smaller elements), generative (e.g., smaller elements can be combined in an unlimited number of ways based on rules of combination), and include dependencies between elements (e.g., individual elements are dependent on, or connected to, other elements within a sequence).

The aim of this thesis is to elucidate some of the complex connections between music and language in relation to syntactic processing. I report on nine experiments and one pilot study that explore syntactic processing in music and language under various conditions. To provide a context for these experiments, Chapter 1 provides an overview of the research to date in this area. First, I discuss the auditory processing of music and speech in relation to their acoustic signals and stages of auditory processing. Second, I provide an overview of syntactic structure in language and music, and discuss some influential theories of syntax. Third, I review existing research on the processing of music and language syntax in the brain. Fourth, I outline two influential theories of shared syntactic processing in music and language: the shared syntactic integration resource hypothesis (SSIRH; Patel, 2003, 2008) and the syntactic equivalence hypothesis (SEH; Koelsch, 2013), and discuss research supporting these theories. Fifth, I discuss some outstanding issues in the syntactic interference literature. Finally, I summarise each of the chapters and experiments that comprise this thesis.

Auditory Processing of Music and Speech

The acoustic signal. Music and spoken language (speech) are auditory signals that involve meaningful changes in pitch, timbre, rhythm, and intensity (Koelsch, 2013; Moreno, 2009; Patel, 2008). *Pitch* is correlated with the fundamental frequency of the sound wave entering the ear, and corresponds to the perception of "highness" and "lowness" of a sound. In music, changes in pitch correspond to melodic intervals; in speech, changes in pitch (speech intonation) are used to convey linguistic stress, or communicate the difference between a question and a statement. *Timbre* is a perceptual attribute of sound that depends on spectral content and amplitude envelope (onset and offset characteristics). Timbre has been defined as the difference in sound quality between two tones when other aspects such as pitch and loudness remain the same (McAdams, 2013). For example, the different sounds created by a piano and a guitar when playing the same note reflect a difference in timbre, and different speakers' voices have different qualities of timbre that distinguish one voice from another. Music and speech also have *rhythm*. Rhythm refers to how sound is organised in time, and is related to grouping, accents, and timing (Patel, 2008). In music, rhythm is often defined as the combination of meter and grouping, whereby meter refers to a predictable and periodic alternation of strong and weak accents or *beats* (Patel, 2008). Whereas musical rhythm tends to be periodic, speech rhythm is one aspect of prosody and tends to relate to the accenting and grouping of certain syllables or words when speaking (Nooteboom, 1997; Patel, 2008). *Intensity* refers to the sound pressure level (SPL) that reaches the eardrum, and is correlated with the perception of loudness. In music and speech, changes in intensity may be introduced to highlight certain events over others (e.g., notes or chords in music; syllables or words in speech).

Although music and speech share a number of acoustic and rhythmic qualities, these qualities can be utilised in different ways. The processing of music and speech rely on *spectral* (frequency based) and *temporal* (timing based) analysis of the incoming auditory signal. These processes are interrelated, but they appear to operate differently depending on whether the incoming signal is music or speech (Farbood, Heeger, Marcus, Hasson, & Lerner, 2015; Patel, 2008; Zatorre, Belin, & Penhune, 2002). In terms of spectral information, speech is confined by the pitch-range of the vocal chords, whereas music has a larger and more varied acoustic signal. Spectral information is usually represented using a spectrogram (see Figure 1). A spectrogram provides a visual representation of changes in frequency and intensity over time, where time is on the x-axis and frequency on the y-axis. Within these axes, intensity is reflected with a grey scale. The processing of speech relies on rapid and fine-grained temporal and spectral information. This fine-grained analysis is necessary for listeners to distinguish between similar speech sounds, such as the phonemes *s* and *f* (Patel, 2008; Zatorre et al., 2002). It is suggested that music processing operates on a longer time scale, as different notes do not occur as rapidly as different phonemes that need to be distinguished (Zatorre et al., 2002). As can be observed in Figure 1a, individual notes can be clearly distinguished within the melody, whereas in 1b, different words are harder to distinguish.

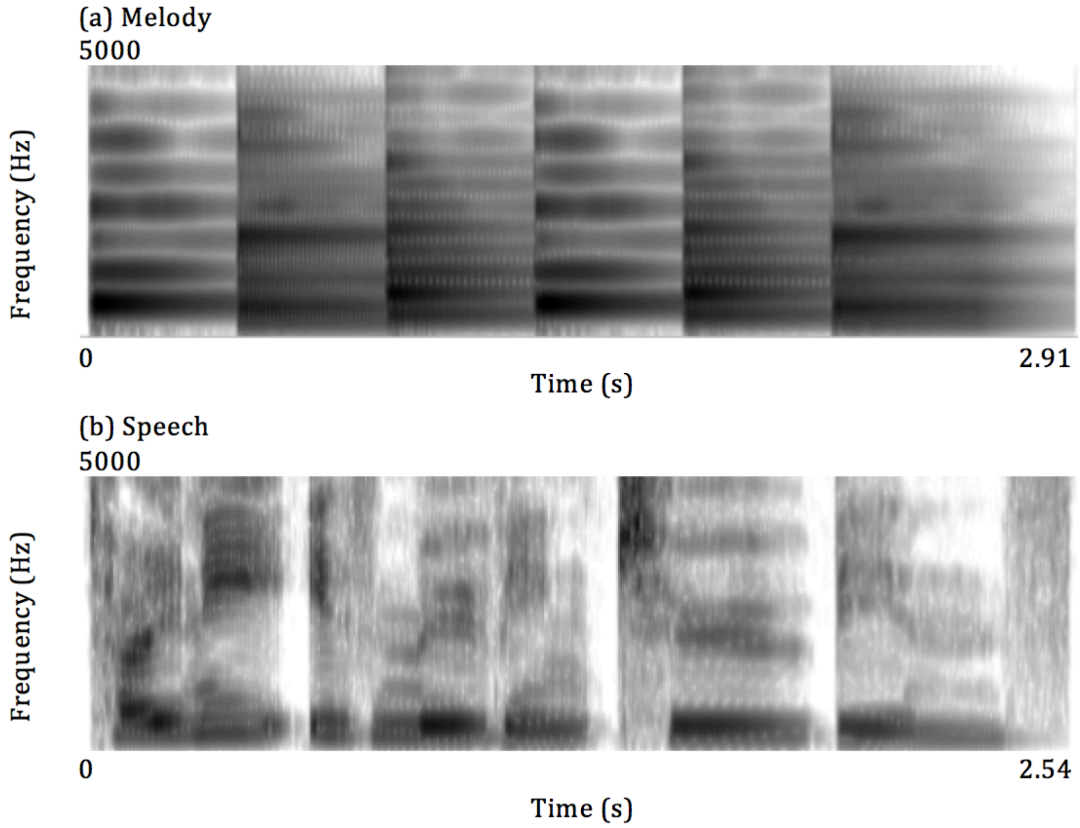

Figure 1. Spectrograms of (a) a six-note melody played with piano, and (b) the sentence *Tony is moving two bins*. S = seconds. Spectrograms generated with Praat (Boersma, 2001).

Zatorre et al. (2002) have suggested that different brain regions are specialised to process spectral and temporal aspects of music and speech. Specifically, the left auditory cortex is more sensitive to temporal information and hence optimal for the analysis of speech, whilst the right auditory cortex is more sensitive to spectral information and hence optimal for the analysis of music. However, there is also evidence that the processing of music and speech is not so distinct. Research suggests that music and speech may be processed in a similar way in infants who do not yet understand the semantic information in language (Brandt, Slevc, & Gebrian, 2012; Trehub & Trainor, 1993). Further, musical cues such as rhythm, melodic contour (the patterns of ups and downs in pitch), and contrasting timbres are thought to scaffold the process of speech acquisition in infants (Brandt et al., 2012). This view is supported by the ubiquitous use of *infant-directed speech* in communicating with young children (Trehub, 2016; Trehub & Trainor, 1993). Infant-directed speech exaggerates the musical aspects of speech, and hence features a larger pitch range and higher pitch than adult-directed speech. It also features simple pitch contours, regular rhythm, and a slow tempo (Brandt et al., 2012; Trehub, 2016; Trehub & Trainor, 1993). Infants exhibit a natural preference for infant-directed speech compared to adult-directed speech (Fernald, 1985). Such links between music and speech early in life suggest an important connection between music and speech processing in the brain.

In adults, there is evidence for cross-domain transfer between music and speech, suggesting common neural substrates (Patel, 2011). Musical training has been shown to enhance language abilities such as speech segmentation (François, Chobert, Besson, & Schön, 2013) and sensitivity to emotional prosody (Thompson, Schellenberg, & Husain, 2004). A range of research has also suggested that musical training enhances pitch perception in speech (Besson, Schon, Moreno, Santos, & Magne, 2007; Bidelman, Gandour, & Krishnan, 2011; Magne, Schon, & Besson, 2006; Marques, Moreno, Castro, & Besson, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007), and speakers of tonal languages have enhanced musical abilities, suggesting that this cross-domain transfer is bidirectional (Bidelman, 2013; Bidelman et al., 2011). Furthermore, the *speech-to-song illusion* (Deutsch, Henthorn, & Lapidis, 2011) demonstrates that a speech signal is perceived melodically after a number of repetitions, suggesting that the distinction between music and speech is not always clear. Such transfer effects and similarities between music and speech suggest commonalities in cognitive processing.

Music and speech are both rhythmic structures that unfold in time. Although music has a more temporally regular and predictable rhythm compared to speech, oscillations in the brain entrain to both music and speech rhythms, and help to encode incoming information (Giraud & Poeppel, 2012; Nozaradan, 2014; Nozaradan, Peretz, Missal, & Mouraux, 2011; Peelle & Davis, 2012; Poeppel, 2014). Importantly, rhythm in music and speech groups incoming auditory information into hierarchical structures (e.g., patterns of strong and weak beats in music, patterns of stressed and non-stressed syllables in speech) and phrase boundaries (e.g., indicating breaks between musical phrases or sentences; Lerdahl & Jackendoff, 1983; Patel, 2006). Research has shown that priming children who have specific language impairment with temporally regular music results in better performance on a language syntax task (Bedoin, Brisseau, Molinier, Roch, & Tillmann, 2016), and there is mounting evidence for a strong link between rhythm skills and grammar skills for typically developing children (Gordon, Jacobs, Schuele, & McAuley, 2015). Similarities have also been observed between rhythmic features of speech and rhythmic qualities of music within languages, suggesting that speech rhythm may influence the composition of classical music (Patel & Daniele, 2003; Patel, Iversen, & Rosenberg, 2006). Thus, music and speech can be compared on a number of levels.

Stages of auditory processing. The stages of music and speech processing are similar across the two domains. The neurocognitive model of music perception (Koelsch, 2011, 2013) and the neurocognitive model of auditory sentence processing (Friederici, 2002) are two current models of music and speech perception that offer a timeline of auditory processing in the brain (see Figure 2 for an outline and comparison of the two models). Koelsch (2013) outlined eight stages of music perception, and Friederici (2002) suggested seven stages of sentence processing within four phases that correspond to separate neural responses. Koelsch (2013) suggested that the first two stages of music perception relate to auditory feature extraction—the translation of incoming acoustic information (e.g., frequency, intensity) into neural activity, and the transformation of this information to cognitive percepts such as pitch and loudness. The third stage, echoic memory and Gestalt formation, is where auditory sensory memory is engaged, and the extracted auditory features begin to be grouped into meaningful percepts based on Gestalt principles of auditory stream segregation (Bregman, 1990). An auditory stream is a coherent group of acoustic information that originates from the same source, and is defined as a "perceptual unit that represents a single happening" (Bregman, 1990, p. 10). For example, an auditory stream could be a single melody or a spoken sentence. The fourth stage—the analysis of intervals—is a more detailed analysis of the auditory stream that encodes relationships between pitches. The fifth stage is structure building, where dependencies between elements are formed based on input, creating phrase-structure and syntax. The sixth stage is structural reanalysis and repair. This stage occurs if the initial parse of the music needs to be revised in some way (e.g., if the music is in a different key than originally thought), and includes larger demands on working memory. Stages seven and eight relate to effects of music on the autonomic and endocrine system (called vitalization), links to premotor processes, and the processing of musical meaning.

A similar timeline is proposed in the neurocognitive model of auditory sentence processing (Friederici, 2002). The initial phase (phase 0) includes primary acoustic analysis, identification of phonemes, and identification of word forms. These stages appear similar to the feature extraction and Gestalt formation stages of the music perception model. Phase 1 includes identifying syntactic information about word category, lexical representations of syntactic information, and syntactic structure building. In phase 2, syntactic and semantic information start to interact. Phase 3 refers to processes of reanalysis and repair, similar to the reanalysis and repair in Koelsch's model of music processing (Koelsch, 2011, 2013). Friederici (2002) also linked these online processes to working memory, whereby phonological memory is associated with phase 0, memory for syntactic structure is associated with phase 1, memory for semantic features and thematic structure is associated with phase 2, and “general” memory resources are associated with phase 3. Working memory therefore appears critical for sentence processing. It should be noted that this model is a syntax-first model, as it implies syntactic information is processed before semantic information, not in parallel. The timeline of syntactic and semantic interaction in the processing of sentences has not been fully clarified, and is still a debate in the literature (Hagoort & Poeppel, 2013). However, for present purposes, the neurocognitive model of auditory sentence processing appears to encompass many of the phases of sentence processing, and so will be used as a comparison to music processing.

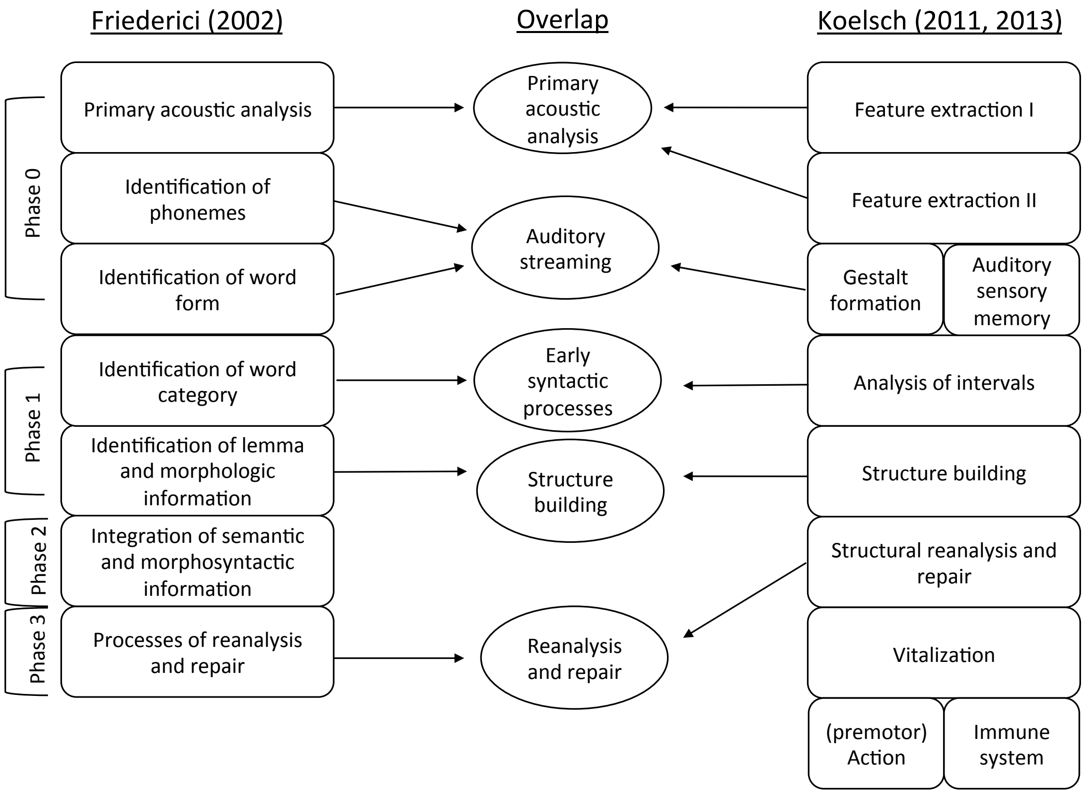

Figure 2. A comparison of the stages in the neurocognitive model of auditory sentence processing (Friederici, 2002) and the neurocognitive model of music perception (Koelsch, 2011, 2013). The middle column shows potential sources of overlap between the two models. Note that the "meaning" and "emotion" stages of the neurocognitive model of music perception are not included in the diagram, as musical meaning and emotion are intricately linked to many stages of music perception, and the link with semantic information in language is less clear.

Syntax in Music and Language

The neurocognitive model of music perception and the neurocognitive model of auditory sentence processing include stages of syntactic structure building and structural reanalysis and repair. It is clear that syntax is integral to the processing of music and speech, and syntax appears to operate in a similar way across the two domains. The term syntax refers to the general principle of combining smaller elements into larger sequences according to a set of combinatorial principles (Patel, 2008). Syntax has been studied in a number of domains such as linguistics (Carnie, 2013; Chomsky, 1957), music (Lerdahl & Jackendoff, 1983), action (Fadiga, Craighero, & D'Ausilio, 2009), and even mathematics (Friedrich & Friederici, 2009). The commonality between all of these domains is that smaller elements (be they words, notes, actions, or numbers) are combined to form larger, hierarchically structured expressions (such as sentences, musical phrases, action sequences, and mathematical formula). Hierarchical structure refers to the idea that smaller elements are grouped within larger elements. This "embedded" structure creates hierarchy within a syntactic sequence, where certain elements are more important and stable than others, and connections and dependencies exist between individual elements.

Language syntax. The study of linguistic syntax as a formal research program seeks to explain linguistic structure through the formulation of principles and rules of combination. This field was pioneered by Noam Chomsky through a number of theories of syntax and syntactic constructions (e.g., Chomsky, 1957). Chomsky (1957, p. 11) first defined syntax as "… the study of the principles and processes by which sentences are constructed in particular languages". The linguistic view of syntax therefore refers to how words (with various word classes) are combined according to syntactic principles to form larger units such as phrases and sentences. Syntax is therefore a level above phonology (the sounds of speech) and morphology (how individual morphemes are combined to form words), and below semantics (the meaning of sentences) and pragmatics (how language is used contextually). To create syntactic structures that communicate meaning, words necessarily belong to different types of word classes (or syntactic categories) that fulfil different functions in a sentence. In English, some of these word classes include verbs, nouns, adjectives, adverbs, and prepositions (Carnie, 2013; Tallerman, 2015). The role of the word within a sentence, and its relation to other words in the sentence, are what defines the word class (Carnie, 2013). Words are grouped into what linguists call constituents—"a group of words that function together as a unit" (Carnie, 2013, p. 73). It is these constituents that form the basis of the syntactic operations that create hierarchical structure.

Linguists represent hierarchical structure within a sentence using syntactic trees. An example of a simple syntactic tree is in Figure 3, adapted from Carnie (2013). This figure shows the constituent structure within the sentence: *The student loved her syntax assignments*. At the highest level, this sentence consists of a simple clause in past tense. Within this clause are determiner phrases (e.g., *the student; her syntax assignments*), a verb phrase (e.g., *loved her syntax assignments*), and noun phrases (e.g., *student; syntax; syntax assignments*). Syntactic trees represent hierarchical structure by showing which elements in the terminal string form constituents, and how those constituents can be embedded inside one another to form larger constituents. The principles of syntax can generate an infinite and unbounded hierarchical structure. However, the processing limitations on working memory capacity constrain the complexity of the hierarchical structures that can be successfully parsed in a sentence.

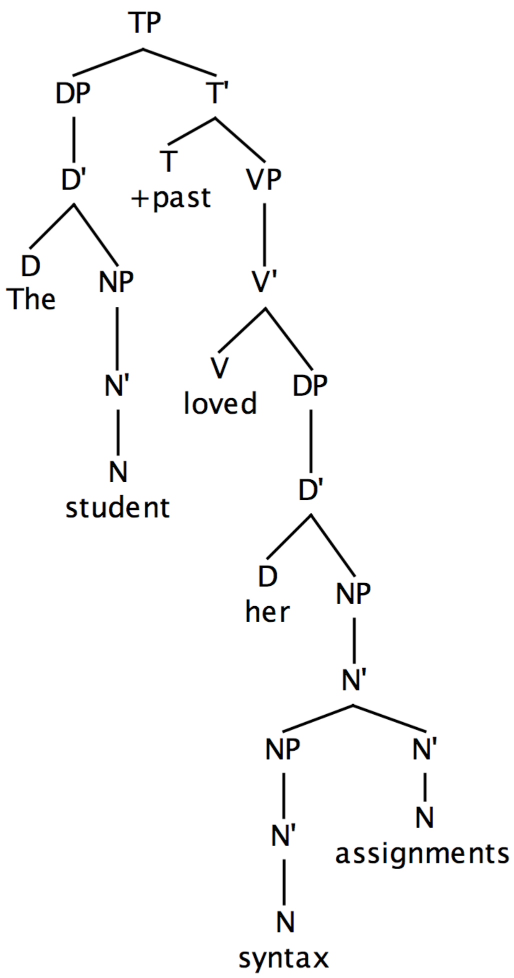

Figure 3. An example of a syntactic tree diagram adapted from Carnie (2013), where TP = tense phrase, DP = determiner phrase, VP = verb phase, NP = noun phrase, D = determiner, T = tense, N = noun, V = verb, and the apostrophe refers to an intermediate level projection.

In addition to describing and understanding the complexities of grammar across multiple languages, linguists are also interested in how children acquire language and learn complex grammar. Many linguists argue that children are born with an innate language faculty, or *universal grammar* that allows them to know which sentences are grammatically acceptable and which are not, even if they have not heard a specific sentence construction previously (Carnie, 2013; Crain & Lillo-Martin, 1999). The theory of universal grammar draws its argument from the logic that infinite, rule based systems are unable to be learned, as children could never be exposed to all possible grammatical combinations of words. Therefore, according to universal grammar, our ability for syntax must be innate, and substantial similarities across different languages of the world is predicted. However, this viewpoint is controversial, and it has been suggested that evidence for a universal grammar is lacking (Dabrowska, 2015).

Given there are approximately 7000 human languages in existence, there must be a cognitive mechanism that is flexible enough to permit a developing child to acquire knowledge of any one of these languages (Levinson, 2013). One such general mechanism is statistical learning (Romberg & Saffran, 2010; Saffran, 2003). Statistical learning refers to the set of processes by which the brain tracks input in the environment in relation to probabilities and co-occurrence of elements—in this case, words (Romberg & Saffran, 2010). Levinson (2013) suggested that the combination of statistical learning and a “language-ready brain” allow for rapid acquisition of syntactic knowledge, and further suggested that our evolved vocal-auditory system and cooperative communication instincts paved the way for the emergence of syntax.

Statistical learning also accounts for the finding that humans are very good at predicting what words or class of word will come next in a sentence (Kuperberg & Jaeger, 2016). Electrophysiological recordings show distinct brain responses reflecting violations of syntactic expectations in language (Friederici, 2002), and numerous studies have shown that the brain predicts upcoming language (see Kuperberg & Jaeger, 2016). However, the role of prediction in language has a shorter history than might be expected. Linguists have conventionally thought prediction to have a small role in sentence processing, as the immense number of possibilities would make it difficult to predict upcoming words (Huettig, 2015). Huettig (2015) argued for the importance of prediction, but also maintained that prediction is not necessary for language processing. However, prediction is an important aspect of language processing that also shares similarities with music (Patel & Morgan, 2016), and will be discussed further in the following section.

Music syntax. The study of syntax is well established in the linguistic literature. In comparison, research on music syntax is relatively new. It has long been known that music has complex hierarchical structure, and a number of attempts have been made to outline and represent this structure (e.g., Agawu, 1989; Lerdahl, 1988; Lerdahl & Jackendoff, 1983; Lerdahl & Krumhansl, 2007; Narmour, 2014; Rohrmeier, 2011). Before discussing how musical structure is formally described, it is useful to outline the elements of music that allow structural hierarchies to exist, as they are distinct from those observed in language. Western tonal music consists of 12 discrete pitches: C, $C^{\#}$, D, E^{b}, E, F, $F^{\#}$, G, $G^{\#}$, A, B^{b}, and B (where $^{\#}$ refers to a sharp, and ^{b} refers to a flat. E^{b} and B^{b} could also be described as $D^{\#}$ and $A^{\#}$ respectively). The distance between each pitch is referred to as an interval, or a *semitone*. Therefore, the distance between E and F is the same as between C and $C^{\#}$. Each pitch can be the *tonic* or base note of a scale that reflects a particular key. For example, the key of C major is made up of the scale notes: C, D, E, F, G, A, B, C. In the C major scale, there are no sharps or flats, and when looking at a piano, these are all the white notes on the keyboard. Within this scale, the C note is the tonic note—the most stable note in the key. The other notes follow in the order: tone, tone, semitone, tone, tone, tone, semitone. A tone is made up of two semitones. Therefore, the distance from C to D is a tone because it includes two semitones (C-$C^{\#}$-D). This pattern of tones and semitones is true for all major keys (e.g., the key of A consists of the scale notes A, B, $C^{\#}$, D, E, $F^{\#}$, $G^{\#}$, A). Each of the 12 scale notes also has a minor key associated with it. For example, each major key has a relative minor that shares all the same notes, but starts on a different tonic note. For example, A minor is the relative minor of C major, which is made up of the scale notes A, B, C, D, E, F, G in the configuration: tone, semitone, tone, tone, semitone, tone, tone.

Crucially for music syntax, some notes within a key are more stable than others, and this difference in tonal stability creates expectations for particular notes or chords. Each key has two other keys that contain all the same scale notes except for one. These keys are the most related. For example, the key of C is most closely related to the keys of G (scale notes G, A, B, C, D, E, $F^{\#}$, G) and F (scale notes F, G, A, B^{b}, C, D, E, F), as they only differ by one note. The relationship between keys can be visualised using the circle of fifths (Figure 4), where each neighbouring major key only differs by one scale note. Within a scale, each note has a scale degree, and certain notes hold a stronger position in relation to the tonic than others. The first note is the first degree, the second note is the second degree, and so on. A C major chord is made up of the tonic (C), the major third (E), and the fifth (G). As such, the major third and the fifth note positions are comparatively stable notes within the underlying scale. Krumhansl (1979) conducted a number of experiments demonstrating that listeners experience notes within a tonal hierarchy based on the key the notes are in. These experiments were the first strong evidence that notes are perceived relative to other notes within the key, and not just based on frequency information. As some words are more important within a sentence, so too are some notes.

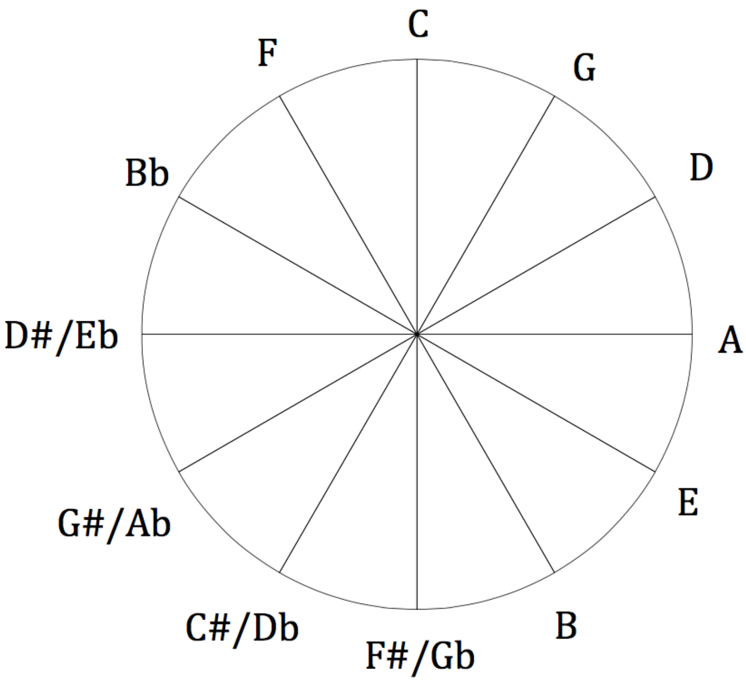

Figure 4. Circle of fifths for major keys. Physically closer keys share more notes than physically distant keys. Figure adapted from Fiveash and Pammer (2014).

In relation to musical stability and tonal relationships, there is a distinction between tonal hierarchies, as described above, and event hierarchies (Bharucha, 1984; Lerdahl, 1988). Tonal hierarchies refer to long-term, abstract representations of pitch relationships based on a culmination of musical experience (Lerdahl, 1988), likely via statistical learning (Jonaitis & Saffran, 2009). Research with infants has shown that tonal hierarchies emerge based on input, and are learned (Trainor & Trehub, 1992). Tonal hierarchies are not structured in time, but are reflected in the perceived pitch relationships observed by Krumhansl (1979). Event hierarchies on the other hand are built in real time when listeners actually hear a piece of music. These hierarchies reflect the relationships between the notes in a piece of music, and draw upon long-term knowledge of tonal hierarchies.

One of the most influential formalisations of music syntax and the psychological implications of musical structure is the generative theory of tonal music (GTTM;

Lerdahl & Jackendoff, 1983). Though recognising the distinctions between music and language syntax, Lerdahl and Jackendoff (1983) created a theory of music syntax that drew inspiration from the linguistic tradition, and created the basis for the scientific study of music syntax. Importantly, they argued that "one should not approach music with any preconceptions that the substance of music theory will look at all like linguistic theory" (Lerdahl & Jackendoff, 1983, p. 5). It is argued by many researchers that trying to equate music and language in terms of their elements (e.g., nouns, verbs, notes, chords etc.) is largely futile, and any important (and useful) comparisons are at a higher, abstract level of combinatorial principles (Jackendoff, 2009; Lerdahl & Jackendoff, 1983; Patel, 2008). Lerdahl and Jackendoff (1983) suggested that musical structure is instead based on pitch and rhythmic organisation, dynamics, timbre, and motivic-thematic elements. Such elements create patterns of tension and resolution, suggested to be integral to musical meaning and emotion (Huron, 2008; Meyer, 1956).

The GTTM outlines four main components of music that are hierarchical in nature. These are grouping structure, metrical structure, time-span reduction, and prolongational reduction (Lerdahl & Jackendoff, 1983). Grouping and metrical structure relate to the way incoming elements are grouped and perceived in time. For example, incoming elements are grouped into hierarchical units (sections, phrases, motives) within a piece, and arranged within a metrical framework of strong and weak beats. Time-span reduction relates to grouping and metrical structure, and is based on the temporal importance of each element within a sequence. Prolongational reduction describes the perception of hierarchical tension and relaxation patterns within the music. These patterns are created through harmonic and melodic aspects of the piece that are perceived in relation to knowledge of tonal hierarchies. A difference between music and language syntax is that language contains clear grammaticality judgements (e.g., language either is or is not grammatically acceptable), whereas the ambiguous nature of music results in *more preferred* and *less preferred* musical structures. To encapsulate the ambiguity of music, Lerdahl and Jackendoff (1983) distinguished between *well-formedness rules* and *preference rules*. Well-formedness rules are based on structural aspects of music, whereas preference rules are based on the psychological preference for some combinations of notes over others. Well-formedness and preference rules draw on the four components mentioned above. Lerdahl and Jackendoff (1983) also added transformational rules to refer to aspects of music structure that were not covered by well-formedness rules.

Extending the GTTM, Lerdahl (1988) added a *tonal pitch space*. The tonal pitch space corresponds to the tonal hierarchies revealed by Krumhansl (1979), and is based on the long-term knowledge of relationships between tones in a mental space. The tonal pitch space relates to so-called stability conditions in the GTTM, and influences how incoming sound sequences are built into event hierarchies (see Figure 5). Lerdahl (1988) expressed the tonal pitch space theory in terms of five levels, where the lowest level refers to each possible note regardless of key, and the highest level reflects octave equivalence (the same note across two octaves, e.g., C in one octave, C in the next octave). Each level increases in tonal importance, with higher levels in the pitch space representing increased structural importance within a key. Behavioural evidence has suggested that melodies containing a hierarchical tonal structure are recalled more accurately than melodies without a hierarchical tonal structure (Deutsch, 1980). These results suggest that tones are represented hierarchically in the brain, and the tonal pitch space theory provides a framework to account for the relationships between pitches within this hierarchy. Extensions and testing of this model showed that prolongational structure, pitch-space, surface-tension, and attraction can explain tonal tension in music

(Lerdahl & Krumhansl, 2007).

Image removed due to copyright, please see citation below

Figure 5. Representation of the generative theory of tonal music (Lerdahl & Jackendoff, 1983), including stability conditions. Figure from Lerdahl (1988).

Another important aspect of music syntax is based on expectation and prediction, and how musical structure relates to tension and resolution patterns in music. Meyer (1956) formalised the connection between musical emotion and expectation in music. He further suggested that musical expectations are based on general Gestalt grouping principles, and are dependent upon memory. Meyer (1956) influenced many subsequent music theorists, including Eugene Narmour (Narmour, 1990, 1992) and David Huron (Huron, 2008). Narmour's implication-realization model is based around expectation, and posits that an unresolved melody (labelled non-closure) affords implications (or creates expectations) for the listener (Narmour, 1992, 2014; Schellenberg, 1996; Thompson, 1996). Like Meyer (1956), Narmour explains these expectations in terms of Gestalt grouping principles such as proximity, similarity, and symmetry. The implications generated by an unresolved melodic fragment can be "realised" to different extents: completely realised, partially realised, partially denied, or completely denied. The model suggests a distinction between bottom-up expectancies that are stimulus driven and largely automatic, and top-down processes that are cognitively learned and more flexible. The ITPRA model proposed by Huron

(2008) considers a more detailed analysis of the forms that expectancy may take, including imagination, tension, prediction, reaction, and appraisal processes. The ITPRA model emphasises the predictive and tension-resolution patterns within musical structures. Links between predictive processes in music and language have also been theorised (Patel & Morgan, 2016).

The Processing of Music and Language Syntax in the Brain

The similarities in music and language syntax have motivated investigation into whether the brain processes syntax in music and language in the same way. The evidence for shared processing between music and language syntax is mixed. In favour of distinct processing resources, a number of neuropsychological studies have reported a double dissociation between music and language, whereby music processing is impaired and language processing is intact, and vice versa. Amusia is a musical disorder (either congenital or acquired) that results in impaired music processing, often relating to both pitch and rhythm dimensions (Ayotte, Peretz, & Hyde, 2002; Sun, Lu, Ho, & Thompson, 2017). Aphasia is a language disorder often acquired after stroke that can impact multiple aspects of language processing, relating to both production and comprehension (Kirshner, 2012).

Pertinent to the current discussion are reports of individuals with amusia who exhibit intact language processing skills (Ayotte et al., 2002; Peretz et al., 1994; Piccirilli et al., 2000), and individuals with aphasia who have intact music processing skills (Basso & Capitani, 1985; Slevc, Faroqi-Shah, Saxena, & Okada, 2016; Tzortzis, Goldblum, Dang, Forette, & Boller, 2000). Such double dissociations provide strong evidence for distinct pathways in the brain for music and language. However, Patel, Iversen, and Hagoort (2004) found that harmonic priming in music was lacking in people with aphasia compared to control participants, and further work has suggested that people with amusia have impaired processing of emotional prosody (Thompson, et al., 2012) and reduced phonological awareness (Sun et al., 2017) in language. Further, many cases of aphasia without amusia are observed in professional musicians, who would be expected to have stronger neural networks for music processing than non-musicians and might not be representative of the general population (Patel, 2008). People with aphasia also benefit from music therapy, and melodic intonation therapy in particular, further supporting a link between the two domains (Belin et al., 1996; Schlaug, Marchina, & Norton, 2008; Wilson, Parsons, & Reutens, 2006).

In contrast to research reporting dissociations between music and language processing, a number of studies have supported the concept of a network of brain areas involved in syntactic processing in music and language. These areas include Broca's area in the left inferior frontal cortex (LIFC), the pars orbitalis in the LIFC, Wernicke's area in the left superior temporal cortex, the superior temporal gyrus (STG), and the posterior medial temporal gyrus, among others (Caplan, 2001; Embick, Marantz, Miyashita, O'Neil, & Sakai, 2000; Friederici, 2002; Grodzinsky & Friederici, 2006; Kemmerer, 2015; Koelsch, Gunter, et al., 2002; Levitin & Menon, 2003). A complication in comparing the processing of music and language syntax is that there is conflicting evidence across various studies about which areas are activated by "syntax". This conflicting evidence occurs because of the difficultly isolating syntax from semantics, working memory, phonology, cognitive control, and so on (Kemmerer, 2015). A recent study has suggested that syntactic processing (in language) is distributed throughout many brain areas, and that investigations should look even more broadly into distributed language networks (Blank, Balewski, Mahowald, & Fedorenko, 2016).

Chapter 1: Introduction

Research has suggested that syntactic processing in music and language involves both frontal and temporal regions of the brain (Hagoort & Poeppel, 2013; Kemmerer, 2015). Broca's area and the inferior frontal gyrus (IFG) are suggested to be involved in integration, working memory, cognitive control and executive functioning; whereas the superior and medial temporal gyri are suggested to hold the actual representations and knowledge that is being accessed (Hagoort & Poeppel, 2013; Kemmerer, 2015). Although Broca's area has long been considered integral to syntactic processing, this view has increasingly been challenged with evidence suggesting that it is instead related to higher-order functions involved in syntactic processing, rather than syntax per se (Brennan et al., 2012; Rogalsky, Rong, Saberi, & Hickok, 2011). Friederici (2002) suggested that Broca's area is likely to be involved in sequencing syntactic input and syntactic memory. This debate is alive and well in the neuroscience of language literature, but in the following section I will focus on studies specifically investigating the processing of music and language syntax.

A number of localisation studies have revealed that the processing of music and language syntax utilises similar neural networks. For example, Maess, Koelsch, Gunter, and Friederici (2001) conducted a magnetoencephalography (MEG) study which revealed that Broca's area (commonly attributed to language syntax processing, see above) and its equivalent in the right hemisphere were sensitive to music syntax. Using functional magnetic resonance imaging (fMRI), Levitin and Menon (2003) further showed that intact music (in comparison to scrambled music), activated the pars orbitalis region of the LIFC, and its equivalent in the right hemisphere. A study by Donnay, Rankin, Lopez-Gonzalez, Jiradejvong, and Limb (2014) had professional jazz musicians improvise and interact (communicate) with another musician while they were in an fMRI scanner. This study revealed that the production of improvised communicative jazz activated language areas compared to when participants were playing but not improvising together. Specifically, Broca's area in the left IFG and the left posterior STG (Wernicke's area) were activated by jazz improvisation. The above studies only investigated the processing of music syntax. However, all reported that areas strongly linked to language processing were activated, especially around Broca's area and the STG. These areas are suggested to be involved in a fronto-temporal network where the frontal areas activate and integrate incoming information based on activated knowledge in the STG (Friederici, 2002; Hagoort & Poeppel, 2013; Patel, 2008). It is suggested that this fronto-temporal network is shared by music and language.

Research examining music and language in the same experiment reveals similar findings. Sammler et al. (2013) used intracranial electroencephalography (EEG) to show that grammatical errors in language and out-of-key chords in music produced the expected early left anterior negativity (ELAN) for language violations and the early right anterior negativity (ERAN) for music violations. These components were localised bilaterally in the STG and in the left IFG. Furthermore, as Sammler et al. (2013) tested the same participants with both the music and language stimuli, they were able to compare the processing of music and language violations within participants to show overlap. Also within-subjects, Kunert, Willems, Casasanto, Patel, and Hagoort (2015) conducted an interference study where they presented complex or basic sentences sung with melodies that were in key, out-of-key, or contained a loudness increase. Kunert et al. (2015) observed increased activation in Broca's area of the IFG that occurred when complex sentences were sung with out-of-key melodies. This activation did not occur with simple sentences or with a loudness increase, and was still present with a subject-by-subject analysis, suggesting an interaction between syntactic processing in music and language. A study by Abrams et al. (2011) also found that both speech and music stimuli activated Broca's area, the pars orbitalis, and left superior and middle temporal gyri. However, a comparison between intact music and speech and reordered music and speech showed that the two domains utilise these resources on a different time scale. These experiments provide further evidence that music and language utilise similar brain networks, though they might be used in different ways.

Although a number of experiments have shown that music and language are processed in similar areas of the brain and elicit similar electrophysiological responses, these findings do not necessarily mean that they share the same neural circuitry (Peretz, Vuvan, Lagrois, & Armony, 2015). Peretz et al. (2015) argued that the density and size of overlapping brain areas would allow for separate neural circuitries within these areas to selectively process music and language. They therefore suggested that the existence of shared neural resources between music and language is still an open question. For example, Rogalsky et al. (2011) presented participants with sentences, scrambled sentences, and melodies. Sentences and melodies activated similar areas in the superior temporal lobe. However, when they isolated hierarchical processes by contrasting sentences with scrambled sentences, and then compared these networks with those activated by melodies, Rogalsky et al. (2011) found that melodies and sentences did not show overlapping activation in the brain. Furthermore, they found that none of their stimuli activated Broca's area. This finding was surprising considering the wealth of previous information showing activation with similar stimuli. The authors suggested that activation was not observed in Broca's area because their stimuli did not contain syntactic violations. However, stimuli without violations have activated Broca's area in the past (e.g., Donnay et al., 2014), so these findings are unclear. Other authors have suggested that overlapping brain areas have only been observed in group analyses, and more detailed analyses should investigate music and language processing within the same individual to clarify whether there is neural overlap (Fedorenko, Duncan, & Kanwisher, 2013; Fedorenko & Kanwisher, 2011). Therefore, more research is necessary to investigate areas of potential overlap in the brain between music and language.

Theories of Shared Syntactic Processing Resources for Music and Language

To recap thus far, there are similarities and differences between music and language in terms of their acoustic signals, stages of processing, syntactic features, and neural processing of syntax. The compelling similarities between music and language in relation to syntactic structure have motivated theories of shared syntactic processing between the two domains. Two such theories are the shared syntactic integration resource hypothesis (SSIRH; Patel, 2008) and the syntactic equivalence hypothesis (SEH; Koelsch, 2013). Both hypotheses predict interference in tasks requiring the simultaneous processing of music and language syntax. Experiments testing the SSIRH and the SEH typically involve the concurrent presentation of music and language, with one or both streams including an out-of-key note or chord, a grammatical error, or heightened ambiguity or complexity (e.g., garden path sentences, which often have to be reanalysed based on new content). The outcomes of these experiments typically suggest that an increased syntactic processing cost in one domain reduces resources available to process syntax in the other domain, resulting in interference. These theories will be discussed in turn.

The SSIRH. The SSIRH (Patel, 2003, 2008) was created to account for the apparent discrepancy in the literature between neuropsychological evidence of dissociation between music and language (e.g., amusia without aphasia and vice versa), and neuroimaging evidence suggesting similarities in processing (e.g., Maess et al., 2001; Patel, 1998), both discussed above. Patel (2008) has suggested that both music and language have *representation* and *resource* networks (see Figure 6). Representation networks hold domain-specific information relevant to music and language, such as harmonic relationships and lexical knowledge. Representation networks can be individually impaired, resulting in situations where individuals can process music but not language (e.g., Slevc et al., 2016). On the other hand, resource networks are domain-general networks that function to process syntax by activating areas in the representation networks. These processing networks are capacity limited, are shared by music and language, and show interference when syntax is processed in both domains simultaneously. Based on music syntax theory, Patel (2008) suggested that an out-of-key musical element has a low level of activation in representation networks, as it is not expected within the tonal context. Therefore, upon encountering an out-of-key element in a tonal sequence, processing networks require more resources to activate the representation in the brain, hence incurring a cognitive integration cost. Patel (2008) suggested that the P600 event-related potential (ERP) component, a positivity that tends to peak at approximately 600 milliseconds (ms), reflects this integrational cost. The P600 appears identical for music and language syntax violations (Patel, Gibson, Ratner, Besson, & Holcomb, 1998). Patel (2008) therefore predicted that interference should be observed in the brain with the concurrent presentation of challenging music and language integrations, as they draw on the same resources.

Image removed due to copyright, please see citation below

Figure 6. The shared syntactic integration resource hypothesis (SSIRH). Music (M) and language (L) share overlapping, domain-general resource networks, but separate, domain-specific representation networks. Reproduced from Patel (2008).

The SEH. Koelsch (2013) proposed a similar theoretical model largely motivated by electrophysiological investigations: the syntactic equivalence hypothesis. The SEH suggests that hierarchically structured domains (e.g., music, language, action, mathematics) share resources related to syntactic processing that are not shared by semantic processes or acoustic deviance processing. Although this theory appears similar to the SSIRH, it encapsulates a larger variety and time range of syntactic processes. The SSIRH suggests shared resources in relation to integrational processes. For example, when an element occurs that is unexpected in the tonal context (e.g., an $F^{\#}$ note in a C major context), the resource networks activate low-activation representations to integrate the unexpected element into the tonal context. This process of integration is reflected electrophysiologically by the P600 ERP component. The SEH also predicts overlap at the level of integration, and further suggests overlap at early stages of initial structure building (approximately 150ms after stimulus onset). Koelsch (2013) suggested that an out-of-key element in a musical sequence interrupts structure-building processes in the brain, and this interruption is reflected electrophysiologically in the ERAN ERP component. Therefore, the SEH also predicts interference when music and language are processed concurrently.

Evidence for Shared Syntactic Processing Between Music and Language

Experimental support for the SSIRH and the SEH is provided by a number of behavioural and ERP experiments. The main method to test syntactic interference is to present music and language concurrently (an interference paradigm) with one or both domains requiring difficult structural integration or interrupting syntactic structure building. To elicit difficult structural integration or interrupt syntactic structure building, syntactic interference experiments to date have all used violations of syntactic structure (e.g., an out-of-key note or chord in music, and a grammatical error in language), an unexpected element (e.g., an unexpected note or chord in music, and a garden path sentence), and/or increases in complexity (e.g., complex sentences compared to more simple sentences).

A study by Fedorenko, Patel, Casasanto, Winawer, and Gibson (2009) combined music and language in sung stimuli. Twelve-word sentences were sung on twelve-note melodies, and each note was sung on a different word (most words were monosyllabic). Sentences were either basic, subject-extracted relative clauses such as: *The boy that helped the girl got an A on the test*, or complex, object-extracted relative clauses such as: *The boy that the girl helped got an A on the test*. The sung melody contained (a) no out-of-key notes, (b) one out-of-key note, or (c) a loudness increase on one note as a control (see Figure 7). Fedorenko et al. (2009) reported that comprehension accuracy was higher in the subject-extracted sentences compared to the object-extracted sentences overall (as expected), and that there was an interaction between music condition (in-key, out-of-key, auditory anomaly) and sentence type. This interaction revealed that the difference between subject- and object-extracted sentences was larger when the melodies were sung with an out-of-key note compared to when they were sung with an in-key note or with an increase in loudness. This experiment was taken as evidence to support the SSIRH, and showed that an out-of-key note reduces comprehension accuracy for complex sentences.

Image removed due to copyright, please see citation below

Figure 7. Example of an experimental trial in Fedorenko et al. (2009). Each word was sung on a different note. Figure from Fedorenko et al. (2009).

Slevc, Rosenberg, and Patel (2009) provided further evidence supporting shared syntactic processing in music and language. Slevc et al. (2009) conducted a self-paced reading task where participants were simultaneously presented with chord sequences and (a) garden path sentences, (b) sentences with a semantically unexpected word, or (c) expected versions of these sentences. Garden path sentences are designed to lead the reader towards one interpretation of the sentence before a disambiguating word occurs that makes the reader reinterpret the sentence in relation to the new information. An example from Slevc et al. (2009) is: *After the trial the attorney advised the defendant* was *likely to commit more crimes*. Such sentences require reanalysis and impose a larger syntactic processing cost than normal sentences (Ni, 1996). An example of a semantically unexpected word in a sentence is: *The boss warned the mailman to watch* *for angry* pigs *when delivering the mail*. A semantically unexpected word does not require a similar reanalysis. These sentences were paired with strongly tonal chord sequences. The harmonic manipulation was an out-of-key chord presented at the same time as the disambiguating word in the garden path sentence, the semantic manipulation in the semantic condition, or at a comparable position in the expected sentences. Slevc et al. (2009) reported slower reading times for the garden path sentences compared to expected sentences when paired with out-of-key chords. This pattern did not occur for the semantic manipulation, or in a control experiment where the out-of-key chord was substituted for a change in timbre. This pattern of results suggests a syntax-specific interference effect.

However, subsequent work has questioned the syntax-specific nature of the finding by Slevc et al. (2009). Perruchet and Poulin-Charronnat (2013) conducted a similar experiment but replaced the syntactic garden path condition with a semantic garden path condition. Semantic garden paths are similar to syntactic garden paths, as the reader initially interprets the sentence one way, but has to re-evaluate that interpretation upon encountering a disambiguating word. Importantly for the garden path sentences in Perruchet and Poulin-Charronnat (2013), the reinterpretation is semantic, not syntactic. An example of a semantic garden path sentence provided by Perruchet and Poulin-Charronnat (2013) is: *The old man went to the bank to withdraw his* net *which was empty*. Upon encountering the word *net*, the sentence has to be re-analysed in terms of a different meaning of the word *bank* (*river bank*). Perruchet and Poulin-Charronnat (2013) argued that unexpected semantic words in the Slevc et al. (2009) stimuli were unable to be integrated into the sentence, and therefore created different demands than the syntactic garden path sentences.

Perruchet and Poulin-Charronnat (2013) observed the same pattern for the semantic garden path stimuli as Slevc et al. (2009) did with syntactic garden paths—the critical word was read more slowly when paired with an out-of-key chord. Perruchet and Poulin-Charronnat (2013) interpreted this finding in a pure attentional framework. They suggested that since the music was not task-relevant, extra resources were only allocated to the music if they were "left over" from sentence processing. As such, Perruchet and Poulin-Charronnat (2013) suggested that semantic errors were more attentionally demanding because they were unpredictable. Further, they suggested that the alternative parse of a garden path sentence could have been pre-activated, and therefore the reanalysis may not have required as much attention as a pure semantic error. Perruchet and Poulin-Charronnat (2013) further argued that because semantic errors required more attentional resources, there were fewer resources available to process the music violation. Therefore, interference was only observed with garden path sentences. However, Slevc and Okada (2014) argued against this interpretation, suggesting that the difference between garden path sentences and sentences with semantic anomalies was that the garden path sentences required revision and reinterpretation. Slevc and Okada (2014) therefore argued that music and language share resources related to cognitive control.

Hoch, Poulin-Charronnat, and Tillmann (2011) investigated the question of shared resources using a lexical decision task, where participants decided if the last word in a sentence was a word or a pseudo-word. Eight-chord sequences were paired with syllable-by-syllable presentation of words. The last chord was either the expected, tonic chord, or a less-expected subdominant chord (the fourth scale degree). The last word of the sentence (when it was a word and not a non-word), was either syntactically expected or unexpected. Sentences were in French, and agreement rules in French syntax require that a gendered article (e.g., masculine or feminine) should be followed by a word with the same gender. Therefore, a gender mismatch creates a syntactic violation. As expected, Hoch et al. (2011) found that syntactically expected words were processed faster than syntactically unexpected words (*the syntactic expectancy effect*), and that words were processed faster in general when they were presented on a tonic chord (a known phenomenon called *tonic facilitation*). Importantly, there was also an interaction: faster processing of the syntactically expected word was reduced when presented with an unexpected chord, and faster processing on the tonic chord disappeared when presented with a syntactically unexpected word. This pattern of results suggest that music and language share resources for processing syntax.

Hoch et al. (2011) conducted a second study where they manipulated semantic expectancy instead of syntactic expectancy. In this experiment, they found that semantically expected words were processed faster than semantically unexpected words (*the semantic expectancy effect*), and that words were processed faster on the expected than the unexpected chord. However, in contrast to the syntactic findings, there was no interaction between musical expectancy and semantic expectancy, suggesting that the semantic expectancy effect was not affected by the tonic facilitation effect and vice versa. These findings suggest shared syntactic processing resources for music and language that are distinct from semantic processing resources.

Fiveash and Pammer (2014) investigated whether interference occurred between music and language in recall for sentences. They presented participants with complex sentences and word-lists paired with music with (a) no violations, (b) a syntactic violation (out-of-key chord), or (c) a timbre violation (chord played with a different instrument). Complex sentences were object-extracted relative clauses such as: *The boy that the dean called to his office had a small voice full of anger*. Word-lists contained five words that were not semantically related and did not contain syntax, such as: *sand, bat, light, pear, mole*. Participants viewed the full sentence or word-list on the screen at the same time as they listened to the music. The syntactic or timbre violations occurred in approximately the middle of each musical sequence so there was time for a tonal context to be built. After stimulus presentation ended, participants recalled the language. Fiveash and Pammer (2014) found that sentence recall was significantly reduced when paired with music with an out-of-key chord. This pattern did not occur for syntax-free word-lists, or with the timbre control, suggesting the results were specific to syntax and were not based on a salient distracting event. This experiment suggests that memory for a complex sentence is impaired when participants concurrently encounter an out-of-key chord in music, supporting theories of shared processing.

Neuroimaging studies have investigated the neural correlates of syntactic interference, and are able provide more clues as to the time frame of syntactic interference. Koelsch, Gunter, Wittfoth, and Sammler (2005) presented chord sequences ending on either an expected, tonic chord, or an unexpected “Neapolitan sixth” chord, at the same time as sentences that were (a) syntactically correct and semantically expected (high cloze probability), (b) syntactically correct and semantically unexpected (low cloze probability), or (c) syntactically incorrect and semantically expected (high cloze probability; see Figure 8). Participants were asked to ignore the music and to focus on the words, and in 10% of trials they had to answer whether the sentence was correct or incorrect.

Image removed due to copyright, please see citation below

Figure 8. Stimuli and design from Koelsch et al. (2005). Chord presentation was aligned with word presentation across all conditions. Figure from Koelsch et al. (2005).

As predicted, (a) unexpected chords paired with expected words elicited an ERAN that peaked at 190ms, (b) syntactic violations in language paired with regular chords elicited a left anterior negativity (LAN) that peaked at approximately 390ms, and (c) semantically unexpected words paired with regular chords showed the expected N400 component. However, when paired with an unexpected chord, the components elicited by syntactic violations and semantically unexpected words were affected differently. The LAN to syntactic violations was significantly reduced when paired with an out-of-key chord, whereas the N400 was not affected (note that this pattern of results was also found by Carrus, Pearce, & Bhattacharya, 2013). Koelsch et al. (2005) suggested this pattern of interactive and non-additive effects was evidence that the processing of syntactic errors in music and language share resources. The authors suggested the LAN was reduced because resources were engaged processing the unexpected chord. The finding that an out-of-key chord did not reduce the response to semantically unexpected words further implies different resources for semantic and syntactic processing in music and language. The ERAN to the unexpected chord was not similarly reduced. Koelsch et al. (2005) suggested the ERAN was not affected because the music was not task-relevant. A control condition showed that the interaction between music and language syntax was related to syntactic processing, and not just a physically deviant element. This study provided important evidence for a neural interaction between syntactic processing in music and language, and showed that this interaction occurs in a very early time window, as predicted by the SEH.

Steinbeis and Koelsch (2008) conducted a similar study to Koelsch et al. (2005). The same conditions of musically expected and unexpected chords were paired with the three language conditions (syntactically correct and semantically expected, syntactically correct and semantically unexpected, syntactically incorrect and semantically expected). However, in this experiment the participants were instructed to attend to both the music and the language. To ensure attention, participants had to detect occasional timbre deviants in the music stimuli, and answer intermittent questions about the language. Similar to the Koelsch et al. (2005) study, syntactic errors in language elicited the LAN and the P600, and semantically unexpected words elicited the N400. The authors also reported a P600 for semantically unexpected words. Replicating the previous study, Steinbeis and Koelsch (2008) observed the ERAN to unexpected chords that peaked at approximately 210ms. In contrast to the Koelsch et al. (2005) study, Steinbeis and Koelsch (2008) observed an additional N5 at approximately 450ms after out-of-key chords. They argued that the N5 reflected the processing of musical meaning, and was elicited because the music in this experiment was task-relevant.

As in Koelsch et al. (2005), Steinbeis and Koelsch (2008) showed that the LAN to the syntactic error in language was reduced when paired with an unexpected chord. However, they also found that the ERAN to the unexpected chord was significantly reduced when paired with a syntactic error in language. Steinbeis and Koelsch (2008) suggested this pattern occurred because participants were asked to attend to both the music and the language. There was no interaction between the N5 elicited by an out-of-key chord and the LAN elicited by a syntactic violation in language. However, there was an interaction between the N5 and the N400 to semantically unexpected words, suggesting that the N5 and N400 use similar resources for processing meaning. The opposite pattern did not occur—the N400 was not reduced by the presence of the N5. These results suggest shared resources for processing structural errors in music and language. Further, they suggest that task-relevant music also elicits a later component related to musical meaning, which interacts with semantic processing in language. This pattern of results could shed light on some of the inconsistent findings between syntactic and semantic interference in the literature.

It is somewhat surprising that Steinbeis and Koelsch (2008) found interference between the N5 to an unexpected chord and the N400 to a semantically unexpected word, as previous research has found no such interaction in dual-task paradigms. Bonnel, Faita, Peretz, and Besson (2001) presented participants with French operatic melodies that contained either semantic anomalies in the sung sentences, melodic anomalies in the melodies, or both anomalies simultaneously. Different participants were asked to detect (a) only semantic anomalies, (b) only melodic anomalies, or (c) both semantic and melodic anomalies. The authors reasoned that if music syntax and language semantic processing used the same resources, an interference effect should be observed in the dual-task that was not evident for the single-task. The results showed no difference in performance between the single-task and dual-task conditions, suggesting distinct resources for processing syntax in music and semantics in language.

The same stimuli were presented in an ERP study to investigate whether music syntax and language semantics shared processing resources (Besson, Faita, Peretz, Bonnel, & Requin, 1998). Besson et al. (1998) predicted that if the processing of semantic incongruities and musical incongruities required distinct resources, there should be an additive effect of brain responses to each type of violation. However, if they drew on the same resources, there should be a dampening effect. Besson et al. (1998) found the N400 to semantic incongruities in language, and what they called the P300 to out-of-key notes in music. In contrast to Steinbeis and Koelsch (2008), Besson et al. (1998) found no decrease in the individual components, but rather an additive effect, suggesting distinct components in the brain for music syntax and language semantic processing. An important difference between the Besson et al. (1998) and Bonnel et al. (2001) stimuli compared to the Steinbeis and Koelsch (2008) stimuli is that the earlier stimuli contained music and language within one auditory stream. The fact that the melody and language were integrated in one auditory stream might trigger different processing compared to chords presented in alignment with written words, with the latter more indicative of interference paradigms. However, a syntactic interference effect was found in sung stimuli in Fedorenko et al. (2009), suggesting that the studies by Besson et al. (1998) and Bonnel et al. (2001) provide important evidence that music syntax and language semantics do not interact in such paradigms, and that an important level of comparison is between music syntax and language syntax.

Outstanding Issues in Syntactic Interference Between Music and Language

There are a number of outstanding issues and gaps in our knowledge in relation to the current evidence for syntactic interference between music and language. In this thesis, three fundamental questions will be addressed which speak to the nature of syntactic processing in music and language in relation to these outstanding issues. One of the primary issues with investigations of syntactic interference to date is that the large majority of experiments have included out-of-key or unexpected elements in music, and syntactic and semantic violations in language. It is unclear what effect these salient and distracting elements have on the concurrent processing of music and language. Further, it is unclear whether the current evidence for syntactic interference between music and language, and the predictions of the SSIRH and the SEH, generalise to music and language without syntactic violations.

The reliance on syntactic violations in the previous literature ties together the three research questions addressed in this thesis. The first question is: can syntactic interference be observed without violations of syntax? If syntactic processing resources are shared between music and language, then the concurrent presentation of intact music and language should also tax syntactic processing resources. In addition to introducing sensory violations that are not purely structural (Tillmann & Bigand, 2015), it is unclear what impact an out-of-key element has on auditory streaming. Out-of-key elements are likely to attract attention, and may interrupt the auditory stream. The connection between auditory streaming, attention, and syntactic processing has rarely been investigated, but appears fundamental to the processing of music and language (Friederici, 2002; Koelsch, 2011). Therefore, the second question addressed in this thesis is: what is the connection between auditory streaming, attention, and syntactic processing, and is this connection similar for music and language? The third question stems from the issue that the majority of studies investigating shared syntactic processing have examined the effect of syntactic violations in music on language rather than vice versa. Because of the additional sensory violations introduced by out-of-key elements, it is not yet clear if the processing of syntax in language interrupts music syntax processing to the same degree that the processing of syntax in music interrupts language syntax processing. Thus, this thesis also investigates the question: can syntactic interference be observed from language to music as well as from music to language?

In addition to these outstanding questions, the underlying neural resources involved in syntactic processing have yet to be elucidated, and are relevant to all three questions. Although it is likely that there are numerous cognitive resources involved in processing music and language syntax, working memory appears to play an important role in the simultaneous processing of the two domains, and its contribution to syntactic processing has rarely been measured in syntactic interference studies. The concept of working memory refers to the processing, storage, and manipulation of information, including attentional processes (Baddeley, 2012; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002). The three main research questions and contributions of working memory will be discussed in turn.

Syntactic violations. One of the confounding issues with the evidence outlined above is that much of it uses violations of syntactic structure in order to investigate shared processing of syntax in music and language. These violations are often necessary to induce challenging structural integration or to disrupt syntactic structure building processes. However, music and language structural violations incur a cognitive cost that may influence the syntactic interference effect. Tillmann and Bigand (2015) argued that out-of-key notes in music violate both syntactic and sensory expectations, as out-of-key notes also introduce harmonics that are not present in the rest of the sequence. These extra psychoacoustic features also influence auditory short-term memory, which is suggested to account for many of the effects of music syntax processing (Bigand, Delbé, Poulin-Charronnat, Leman, & Tillmann, 2014). Some studies of syntactic interference include other "surprising" events (such as changes in timbre or loudness) to try and equate the level of distraction induced by the out-of-key element (e.g., Fedorenko et al., 2009; Fiveash & Pammer, 2014; Slevc et al., 2009). However, timbre changes and increases in loudness do not introduce such strong sensory violations, and therefore may not be adequate control conditions for syntactic violations.

Another method to investigate syntactic interference is to compare the effects of a music manipulation on the processing of syntactic and semantic errors in language. The rationale behind this paradigm is that the concurrent processing of a music syntax manipulation and a language syntax manipulation should produce interference, whereas the concurrent processing of a music syntax manipulation and a language semantic manipulation should not. However, Tillmann and Bigand (2015) have suggested that the comparison of syntactic and semantic violations should be carefully matched, as unexpected semantic words may not be perceived as incorrect as syntactic violations. Semantic manipulations tend to be more ambiguous than syntactic manipulations, as some participants may be able to imagine a number of the scenarios involving low probability semantic sentences. For example, a semantic expectancy manipulation in Koelsch et al. (2005) was: *He sees the cool beer*. This is a perfectly plausible situation, but it is less expected than: *He drinks the cool beer*. These sentences were presented in German, where words are expected to fulfil gender agreements. The syntactic violation in this experiment involved a gender disagreement between the last word *beer* and the preceding information in: *He drinks the cool beer*. As such, the last word was grammatically incorrect, not just unexpected in the context. In the ERP study (Koelsch et al., 2005), separable neural responses were elicited by the syntactic and semantic manipulations. However, behavioural studies are unable to measure cognitive processes at these very early stages, or the depth of processing of semantic compared to syntactic manipulations. Furthermore, most interference paradigms merely assume syntactic and semantic errors are of equal salience, and do not measure participant's detection of errors. It is therefore possible that semantic errors are not as salient as syntactic errors, and this difference could be influencing the syntactic interference effect.

Auditory streaming and attention. In the neurocognitive model of music perception, Koelsch (2013) suggested that auditory streaming processes must occur before syntactic structure can be built. Auditory streaming therefore appears critical to syntactic processing, and yet relatively little research has investigated connections between the two processes. For example, does auditory streaming occur regardless of information about syntactic structure, or is there an interdependent relationship whereby syntactic processing also influences auditory streaming in a developing syntactic representation? It is unclear whether out-of-key elements, timbre changes, and loudness changes (for example) are perceived as belonging to the same auditory stream as the rest of the sequence. It is also unclear whether there is a distinction between stimuli presented within one auditory stream (e.g., a sentence sung on a melody) compared to the concurrent presentation of written sentences and auditory musical sequences. Thus, connections between auditory streaming and syntactic processing remain to be elucidated.

Auditory streaming appears intricately related to attention. Bregman (1990) distinguished between *primitive* auditory streaming and *schema-based* auditory streaming. Primitive auditory streaming occurs pre-attentively and without conscious attention, whereas schema-based streaming is influenced by attention and expectation. Such a distinction would suggest that although auditory streaming occurs without conscious attention, attention does affect auditory streaming processes. In line with this interpretation, research has shown that syntactic processing of music and speech occurs without direct attention, but that attending to music or speech enhances the degree of syntactic processing (Loui, Grent-'t-Jong, Torpey, & Woldorff, 2005; Maidhof & Koelsch, 2011). Furthermore, Maidhof and Koelsch (2011) showed that the brain response to an unexpected chord in music was affected by attention more than the brain response to a syntactic violation in speech. The authors suggested speech was less affected by attention because speech was a more common stimulus than music for their sample of non-musicians. This result could also reflect the possibility that music and speech hold different levels of importance in every day processing, and that speech is often more task-relevant than music in a dual-stimulus situation. This imbalance relates to predictions of whether interference can be observed from language to music.

Effects of language syntax on music syntax processing. Only two studies to date have investigated syntactic interference in the opposite direction than is usually studied: from language to music (Kunert, Willems, & Hagoort, 2016; Van de Cavey, Severens, & Hartsuiker, 2017). Kunert et al. (2016) showed that harmonic closure judgements in music were reduced when the music was simultaneously presented with syntactic garden path sentences compared to semantic garden path sentences and non-anomalous sentences. Van de Cavey et al. (2017) observed that the processing of phrase-boundaries in music were reduced when participants were simultaneously reading syntactic garden path sentences compared to sentences with a syntactic violation or sentences with no violation. These studies provide preliminary evidence that syntactic interference might occur in both directions. Testing whether syntactic interference can be observed in music processing is important to investigate whether theories of shared processing extend in both directions. For example, it is possible that music syntax affects language syntax processing more than language syntax affects music syntax processing. This possible asymmetry cannot be ascertained from the existing literature, as the majority of previous literature has only measured performance on language tasks. However, a shared processing account would predict that interference should be observed in both directions. It is therefore important to test these predictions to further our understanding of the conditions under which syntactic interference occurs, and whether shared processing resources are symmetrical.

Working memory and syntactic processing. The three main research questions investigated in this thesis revolve around the concurrent processing of syntactic structure in music and language. However, the nature of syntactic processing resources shared by music and language is unclear. Music and language processing rely on cognitive resources involving maintenance, integration and manipulation of incoming information, and the tracking of long distance dependencies (Hagoort & Poeppel, 2013; Patel, 2008). These processes appear strongly linked to working memory (Burunat, Alluri, Toiviainen, Numminen, & Brattico, 2014; Fiebach, Schlesewsky, & Friederici, 2002; King & Just, 1991), and it is likely that the syntactic processing of music and language includes some contribution from working memory (Burunat et al., 2014; Kljajević, 2010; Vos, Gunter, Kolk, & Mulder, 2001). Patel (2008) suggested that *structural integration* processes are shared between music and language. Structural integration refers to the process by which incoming information (e.g., a note or a word) is integrated into a developing syntactic representation (e.g., a melody or a sentence), and could be suggested to rely heavily on working memory processes (Patel, 1998). It has also been suggested that the processing of music and language engages cognitive control resources. For example, Slevc and Okada (2014) argued that the concurrent processing of out-of-key notes and garden path sentences is likely to engage cognitive control processes, and hence produce interference. However, it is unclear whether cognitive control resources are engaged under normal listening conditions, or whether they are only engaged when there is some kind of conflict or event that needs resolving. As such, I will focus on working memory as the most likely basis of shared processing between music and language. [1]

Experiments involving the concurrent presentation of music and language are likely to draw heavily on working memory. The central executive, phonological loop, and episodic buffer (Baddeley, 2000) components of the WM model proposed by Baddeley and Hitch (1974) appear to account for a large amount of shared processing between music and language. The central executive is suggested to coordinate four main functions: focusing attention, dividing attention, task switching, and interfacing with long-term memory (LTM). The phonological loop is considered a slave-system of the central executive and stores short-term verbal-acoustic input, and the episodic buffer is suggested to bind information from multiple modalities with representations from LTM. Although the Baddeley and Hitch (1974) WM model does not focus specifically on syntax in sentence or music processing, models of language processing (and the SSIRH) tend to suggest that syntactic processing includes maintenance, online structural integration processes, and the retrieval of specific knowledge from LTM. These aspects could be explained by the phonological loop, episodic buffer and central executive components of the model.

[1] Note that Fedorenko, Behr, & Kanwisher (2011) made a distinction between language processing and more general resources such as working memory. This paper and further motivations for measuring working memory capacity will be explored in Chapter 4.

One model of language processing that relates to working memory resources is the memory, unification, and control (MUC) model (Hagoort, 2013; Hagoort & Poeppel, 2013). In this model, language processing is based on the retrieval and *unification* of long-term information into new syntactic structures. This model aligns well with experimental findings regarding syntactic processing in the brain, and Hagoort and Poeppel (2013) have suggested that the MUC model can also be applied to music processing. In the MUC model, memory processes refer to long-term representations of information (such as words and their syntactic structures) that are held in, and retrieved from, the temporal cortex. Unification refers to the combination of these words online in a *unification space*, largely activating Broca's area in the LIFC. The unification space appears to involve processes of syntactic integration and working memory. The final element, *executive control*, appears similar to the central executive in Baddeley and Hitch (1974), and cognitive control (Cooper, 2010), as it involves attending to relevant information, turn taking in conversations, and other control mechanisms (Hagoort, 2013). The unification and retrieval processes described in the MUC model are likely to be closely linked to the resource networks and shared processing resources used by music and language, and are likely to strongly rely on working memory processes. Thus, working memory appears critical to the processing of music and language syntax, but is not represented in tests of shared processing between music and language (the SSIRH and the SEH).

The Current Thesis

The current thesis aimed to elucidate the nature of shared syntactic processing between music and language by investigating three fundamental questions. First, can syntactic interference be observed in stimuli without violations of syntactic structure? Second, does disrupting auditory streaming processes influence syntactic processing, and is the connection between auditory streaming and syntactic processing similar for music and language? And third, can syntactic interference be observed in music syntax processing as well as language syntax processing? These questions are important to the understanding of how music and language syntax are processed in the brain. To assess the contribution of working memory to syntactic processing, working memory capacity (WMC) was measured and included in analyses. It was hypothesised that WMC would be closely tied to syntactic processing, such that greater WMC may facilitate syntactic processing. Indeed, it has been suggested that performance on tasks used to engage syntactic processing might reflect differences in working memory capacity (e.g., King & Just, 1991). However, the role of WMC in syntactic processing is not developed in either the SSIRH or the SEH, and the relative contribution of WMC to syntactic processing across music and language is unknown. Nine experiments are presented across four experimental chapters that investigate the nature of syntactic processing in music and language.

Chapter 2. Chapter 2 investigated Question 1 (whether interference can occur in music and language without violations of syntax) and Question 2 (whether a disrupted auditory stream influences syntactic processing). Three experiments are reported. Experiment 1 tested the hypothesis that syntactic interference could occur in stimuli without violations of syntactic structure. Participants read complex sentences (with syntax) or word-lists (without syntax) while listening to four different types of auditory stimuli that were manipulated for syntactic structure and level of distraction. The dependent variable was language recall, and there was no music task. As predicted, Experiment 1 showed interference when syntactic musical sequences were paired with syntactic sentences, suggesting that interference can be observed without violations. However, there were also some unexpected results. One of the auditory conditions that had three, alternating timbres playing the melody (to be both syntactic and distracting) resulted in fewer errors in sentence recall than an auditory condition in which the same melodies were played with only one timbre. This result was surprising, considering additive effects of syntactic interference and distraction were predicted. It was hypothesised that the alternating timbres disrupted auditory streaming processes (Bregman, 1990). This disruption then resulted in a less coherent syntactic representation, reduced processing of musical syntax, and therefore decreased interference with sentence processing. This hypothesis was tested in Experiment 2, in which participants compared two melodies that were either played with one timbre, or with three timbres. Experiment 2 revealed that participants were significantly more sensitive to changes in one-timbre melodies compared to three-timbre melodies, supporting the hypothesis that alternating timbres reduce syntactic processing.

In Experiment 3, the role of attention and auditory streaming in syntactic interference was investigated. In this experiment, participants read complex sentences, basic sentences, and word-lists while either attending to a melody in one ear, or an environmental sound sequence in the other ear. The auditory stimuli always included both a melody and environmental sounds—the difference was which stream participants attended to. There was no difference in word-list recall depending on where attention was directed; however, participants were significantly worse at recalling sentences when their attention was directed to the melodic stream compared to the environmental sound stream. Furthermore, detection of out-of-key notes in the melodies was significantly worse when participants were concurrently reading sentences compared to word-lists, a finding that did not occur for detection of gongs in the environmental sequences. Chapter 2 therefore provides evidence for syntactic interference between music and language without syntactic violations, suggests that auditory streaming disrupts syntactic processing, and shows the importance of attention in syntactic processing.

Chapter 3. The aim of Chapter 3 was to further investigate Question 2 (whether disrupting auditory streams reduces syntactic processing). To this end, ERPs were used to measure brain activity to syntactic violations in melodies that contained one timbre compared to three timbres. This paradigm was extended to language to investigate whether structural violations in sentences were noticed more when sentences were spoken by one speaker compared to three speakers. Chapter 3 showed that the brain response to an out-of-key note in a melody (the ERAN) was significantly reduced when the melody contained three-timbres compared to one-timbre. This finding supports the hypothesis that alternating timbres in a melody results in reduced syntactic processing. Phrase-structure violations in language elicited the LAN, which was visually reduced in the three-timbre condition compared to the one-timbre condition; however, this reduction did not reach significance. Chapter 3 demonstrates that when alternating timbres disrupt an auditory stream, syntactic processing is reduced. This phenomena appears to operate similarly in music and language. However, further investigations into language are necessary.

Chapter 4. Chapter 4 continued investigations into Question 1 (whether syntactic interference can be observed without syntactic violations) by manipulating complexity in a dual-task paradigm. In Experiments 1 and 2, complexity was manipulated in auditory stimuli (musical and environmental) and these stimuli were presented at the same time as participants were concurrently performing a language comprehension task (syntactic) or a visuospatial search task (non-syntactic). Increases in musical complexity introduce syntax, whereas increases in environmental complexity do not. It was therefore predicted that interference would be observed for language comprehension when participants were concurrently listening to complex music compared to complex environmental sounds. In Experiment 1, participants were also asked whether the auditory stimulus was musical or environmental to ensure they were paying attention. In Experiment 2, participants were asked to make a more difficult judgement—whether the auditory stimulus was basic or complex. Neither Experiments 1 nor 2 showed a difference in language comprehension or visuospatial search depending on concurrent auditory condition. However, in Experiment 2, interference was observed in the complexity judgements for the secondary task. While performing the visuospatial search task, participants were significantly better at judging complexity of the music stimuli compared to the environmental sound stimuli, suggesting that judging musical complexity was an easier task. However, while engaged in the language comprehension task, there was no difference in complexity judgements for the music and environmental sounds, suggesting that the language comprehension task eliminated the difference between the music and environmental stimuli. These results will be discussed in a model presented in Chapter 6 that incorporates task prioritisation.

Chapter 5. In Chapter 5, the syntactic interference effect was investigated in the opposite direction by measuring effects of language syntax violations on music syntax processing (Question 3: can syntactic interference be observed in music processing?). Experiment 1 was a melodic same-different task, where participants indicated whether two short melodies were the same or different. The first melody was played at the same time as participants read sentences with (a) no error, (b) a semantic error, or (c) a syntactic error. The second melody was presented without a concurrent sentence. Participants first had to indicate whether the melodies were the same or different. They then indicated whether there was an error in the sentence, and finally, what the last word in the sentence was. Experiment 1 showed no differences in same-different judgements depending on sentence condition. However, there were some interesting correlations with WMC, and semantic errors in the language were noticed significantly less than syntactic errors.

In Experiment 2, language to music interference was examined by investigating melody recall when melodies were concurrently presented with spoken sentences (the opposite design to Experiment 1 in Chapter 2). Participants listened to six-note melodies at the same time as auditory sentences that had (a) no error, (b) a semantic error, or (c) a syntactic error. After the melody and sentence were presented concurrently, participants recalled the melody, and then indicated whether there was an error in the language. An analysis of total recall across the whole sequence showed that melody recall was significantly impaired when sentences were presented at the same time as melodies; however, there was no difference depending on sentence condition. A closer examination of the final two notes (where the errors were presented), revealed a clear difference in recall for melodies depending on the concurrent sentence condition. When paired with the semantic error and no error sentences, melody recall *decreased* from the penultimate to final position. However, when melodies were presented on their own or paired with syntactic error sentences, melody recall *increased* from the penultimate to the final condition. As in Experiment 1, semantic errors in language were detected significantly less than syntactic errors and the correct identification of no error.

The results from Experiments 1 and 2 in Chapter 5 suggest that semantic errors in language may not be as noticeable as syntactic errors in language. This finding could mean that semantic errors are not a good control for syntactic errors, as suggested by Tillmann and Bigand (2015). However, if semantic errors are noticed to the same extent as syntactic errors when they are not paired with music, then it is possible that the concurrent music task was affecting semantic error detection in language. To test these possibilities, the written sentences from Experiment 1 and the spoken sentences from Experiment 2 were presented to participants in a single-task paradigm in Experiment 3. Results suggested a distinction between the two types of stimuli. For the written sentences from Osterhout and Nicol (1999), semantic errors were detected worse than syntactic errors in a single-task. This finding suggests that semantic errors may not be an appropriate control for syntactic errors in music-language interference tasks. However, there was no difference between syntactic and semantic error detection in the spoken sentences from Sun, Lu, Ho, Johnson, and Thompson (2015). This finding suggests that semantic errors in the spoken sentence stimuli were only noticed worse than syntactic errors when paired with the melody recall task.

Chapter 6. Chapter 6 summarises all of the findings from Chapters 1 to 5, and discusses how they enhance and extend the current literature. The competitive attention and prioritisation model is introduced to account for the pattern of results observed throughout the thesis, and to provide a broad cognitive framework within which research on syntactic processing may be understood and interpreted. The CAP model highlights two important and interrelated factors that determine whether syntactic interference will occur in a dual-stimulus or dual-task situation. These are task prioritisation (related to depth of processing) and field of attention (related to breadth of processing). Task prioritisation and field of attention are directly influenced by bottom-up processes and stimulus characteristics, as well as by top-down processes and task goals. In a dual-stimulus or dual-task situation, I suggest that participants will allocate more resources to tasks that are given top-down priority (e.g., focus on language, focus on music), or have bottom-up stimulus characteristics that require more attention to process (e.g., complex sentences or demanding pieces of music). Field of attention refers to how attention is allocated across the sequence. An extended field of attention is engaged when the whole sequence must be taken into account (e.g., read a whole sentence, listen to a whole melody). A localised field of attention is engaged when elements of the sequence must be examined individually, rather than as a whole (e.g., detect an error in the language or music), or when a salient event occurs such as an out-of-key note.

The experiments reviewed in the discussion suggest that dual-task requirements place different demands on processing than single-task requirements, and that a cognitive trade-off may occur when completing two tasks simultaneously. Measuring processing of both stimuli in a dual-stimulus experiment is therefore encouraged. Chapter 6 concludes with a discussion of future directions for the CAP model, and some suggestions for future research.

Chapter 1: Introduction

References

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex, 21*(7), 1507-1518. doi:10.1093/cercor/bhq198

Agawu, V. K. (1989). Schenkerian notation in theory and practice. *Music Analysis, 8*(3), 275-301.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125*, 238-251.

Baddeley, A. (2000). The episodic buffer: A new component of working memory. *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29. doi: 10.1146/annurev-psych-120710-100422

Baddeley, A., & Hitch, G. (1974). Working memory. In H. B. Gordon (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Basso, A., & Capitani, E. (1985). Spared musical abilities in a conductor with global aphasia and ideomotor apraxia. *Journal of Neurology, Neurosurgery & Psychiatry, 48*(5), 407-412. doi:10.1136/jnnp.48.5.407

Bedoin, N., Brisseau, L., Molinier, P., Roch, D., & Tillmann, B. (2016). Temporally regular musical primes facilitate subsequent syntax processing in children with specific language impairment. *Frontiers in Neuroscience, 10*(245). doi:10.3389/fnins.2016.00245

Chapter 1: Introduction

Belin, P., van Eeckhout, P., Zilbovicius, M., Remy, P., Francois, C., Guillaume, S., . . . Samson, Y. (1996). Recovery from nonfluent aphasia after melodic intonation therapy: A PET study. *Neurology, 47*, 1504:1511.

Besson, M., Faita, F., Peretz, I., Bonnel, A., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science, 9*(6), 494-498.

Besson, M., Schon, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience, 25*(3-4), 399-410.

Bharucha, J. J. (1984). Event hierarchies, tonal hierarchies, and assimilation: A reply to Deutsch and Dowling. *Journal of Experimental Psychology: General, 113*(3), 421-425. doi:10.1037/0096-3445.113.3.421

Bidelman, G. M. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS One, 8*(4), e60676. doi:10.1371/journal.pone.0060676

Bidelman, G. M., Gandour, J., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23*(2), 425-434.

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience, 8*. doi:10.3389/fnsys.2014.00094

Blank, I., Balewski, Z., Mahowald, K., & Fedorenko, E. (2016). Syntactic processing is distributed across the language system. *NeuroImage, 127*, 307-323. doi:10.1016/j.neuroimage.2015.11.069

Chapter 1: Introduction

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-347.

Bonnel, A., 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. doi:10.3758/BF03194534

Brandt, A. K., Slevc, R., & Gebrian, M. (2012). Music and early language acquisition. *Frontiers in Psychology, 3*. doi:10.3389/fpsyg.2012.00327

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language, 120*(2), 163-173. doi:10.1016/j.bandl.2010.04.002

Brown, S. (2000). The musilanguage model of music evolution. In N. L. Wallin, S. Brown, & B. Merker (Eds.), *The origins of music*. Cambridge, MA: MIT Press.

Burunat, I., Alluri, V., Toiviainen, P., Numminen, J., & Brattico, E. (2014). Dynamics of brain activity underlying working memory for music in a naturalistic condition. *Cortex, 57*, 254-269. doi: 10.1016/j.cortex.2014.04.012

Caplan, D. (2001). Functional neuroimaging studies of syntactic processing. *Journal of Psycholinguistic Research, 30*(3), 297-320.

Carnie, A. (2013). *Syntax: A generative introduction*. UK: Wiley.

Carrus, E., Pearce, M. T., & Bhattacharya, J. (2013). Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. *Cortex, 49*(8), 2186-2200. doi:10.1016/j.cortex.2012.08.024

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chapter 1: Introduction

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163-183.

Cooper, R. P. (2010). Cognitive control: Componential or emergent? *Topics in Cognitive Science, 2*(4), 598-613. doi:10.1111/j.1756-8765.2010.01110.x

Crain, S., & Lillo-Martin, D. (1999). *An introduction to linguistic theory and language acquisition*. Massachusetts, USA: Blackwell Publishers.

Cross, I. (2014). Music and communication in music psychology. *Psychology of Music, 42*(6), 809-819. doi:10.1177/0305735614543968

Dabrowska, E. (2015). What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.00852

Darwin, C. (1882). *The descent of man, and selection in relation to sex* (2nd, Revised and Augmented ed.). London: John Murray.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics, 28*(5), 381-389. doi:10.3758/bf03204881

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *The Journal of the Acoustical Society of America, 129*(4), 2245-2252. doi:10.1121/1.3562174

Donnay, G. F., Rankin, S. K., Lopez-Gonzalez, M., Jiradejvong, P., & Limb, C. J. (2014). Neural substrates of interactive musical improvisation: An FMRI study of 'trading fours' in jazz. *PLoS One, 9*(2), e88665. doi:10.1371/journal.pone.0088665

Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., & Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proceedings of the National Academy of Sciences, 97*(11), 6150-6154. doi:10.1073/pnas.100098897

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences, 1169*, 448-458. doi:10.1111/j.1749-6632.2009.04582.x

Farbood, M. M., Heeger, D., Marcus, G., Hasson, U., & Lerner, Y. (2015). The neural processing of hierarchical structure in music and speech at different timescales. *Frontiers in Neuroscience, 9*. doi:10.3389/fnins.2015.00157

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences, 108*(39), 16428-16433. doi:10.1073/pnas.1112937108

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences, 110*(41), 16616-16621. doi:10.1073/pnas.1315235110

Fedorenko, E., & Kanwisher, N. (2011). Some regions within Broca's area do respond more strongly to sentences than to linguistically degraded stimuli: A comment on Rogalsky and Hickok. *Journal of Cognitive Neuroscience, 23*(10), 2632-2635. doi:10.1162/jocn_a_00043

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition, 37*(1), 1-9. doi:10.3758/MC.37.1.1

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*(2), 181-195. doi:10.1016/S0163-6383(85)80005-9

Fiebach, C., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research, 30*(3), 321-338.

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music, 42*(2), 190-209. doi:10.1177/0305735612463949

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex, 23*(9), 2038-2043. doi:10.1093/cercor/bhs180

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84. doi:10.1016/S1364-6613(00)01839-8

Friedrich, R., & Friederici, A. D. (2009). Mathematical logic in the human brain: Syntax. *PLoS One, 4*(5), e5599. doi:10.1371/journal.pone.0005599

Giraud, A., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511-517. doi:10.1038/nn.3063

Gordon, R. L., Jacobs, M. S., Schuele, C. M., & McAuley, J. D. (2015). Perspectives on the rhythm–grammar link and its implications for typical and atypical language development. *Annals of the New York Academy of Sciences, 1337*(1), 16-25. doi:10.1111/nyas.12683

Grodzinsky, Y., & Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology, 16*(2), 240-246. doi:10.1016/j.conb.2006.03.007

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology, 4*(416). doi:10.3389/fpsyg.2013.00416

Hagoort, P., & Poeppel, D. (2013). The infrastructure of the language-ready brain. In M. Arbib (Ed.), *Language, music, and the brain: A mysterious relationship*. Cambridge, MA: MIT Press.

Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: Syntactic and semantic structures. *Frontiers in Psychology, 2*, 112. doi:10.3389/fpsyg.2011.00112

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research, 1626*, 118-135. doi:10.1016/j.brainres.2015.02.014

Huron, D. (2008). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception: An Interdisciplinary Journal, 26*(3), 195-204. doi:10.1525/mp.2009.26.3.195

Jonaitis, E. M., & Saffran, J. R. (2009). Learning harmony: The role of serial statistics. *Cognitive Science, 33*(5), 951-968. doi:10.1111/j.1551-6709.2009.01036.x

Kemmerer, D. (2015). *Cognitive neuroscience of language*. New York: Psychology Press.

King, J., & Just, M. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*(5), 580-602.

Kirshner, H. S. (2012). Aphasia. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior: Volume one. A-D* (2nd ed.). UK: Elsevier Science.

Kljajević, V. (2010). Is syntactic working memory language specific? *Psihologija, 43*(1), 85-101. doi: 10.2298/psi1001085k

Koelsch, S. (2011). Towards a neural basis of music perception - A review and updated model. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00110

Chapter 1: Introduction

Koelsch, S. (2013). *Brain and music*. Oxford, UK: John Wiley & Sons.

Koelsch, S., Gunter, T., v. Cramon, D., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage, 17*(2), 956-966. doi:10.1006/nimg.2002.1154

Koelsch, S., Gunter, T., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and music: An ERP study. *Journal of Cognitive Neuroscience, 17*(10), 1565-1577.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11*(3), 346-374. doi:10.1016/0010-0285(79)90016-1

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS One, 10*(11), e0141069. doi:10.1371/journal.pone.0141069

Kunert, R., Willems, R. M., & Hagoort, P. (2016). Language influences music harmony perception: Effects of shared syntactic integration resources beyond attention. *Royal Society Open Science, 3*(2). doi:10.1098/rsos.150685

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32-59. doi:10.1080/23273798.2015.1102299

Lerdahl, F. (1988). Tonal pitch space. *Music Perception, 5*(3), 35.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception, 24*(4), 329-366. doi:10.1525/Mp.2007.24.4.329

Chapter 1: Introduction

Levinson, S. C. (2013). Cross-cultural universals and communication structures. In M. Arbib (Ed.), *Language, music, and the brain: A mysterious relationship*. Cambridge, MA: MIT Press.

Levitin, D. J., & Menon, V. (2003). Musical structure procesed in "language" areas of the brain: A possible role for Brodmann area 47 in temporal coherence. *NeuroImage, 20*, 2142-2152. doi:10.1016/S1053-8119(03)00482-8

Loui, P., Grent-'t-Jong, T., Torpey, D., & Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Cognitive Brain Research, 25*(3), 678-687. doi:10.1016/j.cogbrainres.2005.08.019

Maess, B., Koelsch, S., Gunter, T., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4*(5), 540-545.

Magne, C., Schon, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience, 18*(2), 199-211.

Maidhof, C., & Koelsch, S. (2011). Effects of selective attention on syntax processing in music and language. *Journal of Cognitive Neuroscience, 23*(9), 2252-2267.

Marques, C., Moreno, S., Castro, S., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 19*(9), 1453-1463.

McAdams, S. (2013). Musical timbre perception. In D. Deutsch (Ed.), *Psychology of Music*. USA: Elsevier, Inc.

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

Chapter 1: Introduction

Mithen, S. (2007). *The singing neanderthals*. Cambridge, MA: Harvard University Press.

Mithen, S. (2009). The music instinct: The evolutionary basis of musicality. *Annals of the New York Academy of Sciences, 1169*, 3-12. doi:10.1111/j.1749-6632.2009.04590.x

Moreno, S. (2009). Can music influence language and cognition? *Contemporary Music Review, 28*(3), 329-345. doi:10.1080/07494460903404410

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity*. Chicago: University of Chicago Press.

Narmour, E. (2014). Implication-realization. In W. F. Thompson (Ed.), *Music in the social and behavioural sciences: An encycopledia* (pp. 589-593). Thousand Oaks, CA: SAGE Publications, Inc.

Ni, W. (1996). Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes, 11*(3), 283-334. doi:10.1080/016909696387196

Nooteboom, S. (1997). The prosody of speech: Melody and rhythm. In W. J. Hardcastle & J Laver (Eds.), *The Handbook of Phonetic Sciences*. Oxford: Blackwell.

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequence-tagging. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*.

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience, 31*(28), 10234-10240. doi:10.1523/jneurosci.0411-11.2011

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*(3), 283-317. doi:10.1080/016909699386310

Patel, A. D. (1998). Syntactic processing in language and music: Different cognitive operations, similar neural resources? *Music Perception: An Interdisciplinary Journal, 16*(1), 27-42.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674-681.

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception: An Interdisciplinary Journal, 24*(1), 99-104. doi:10.1525/mp.2006.24.1.99

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

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2*, 142. doi: 10.3389/fpsyg.2011.00142

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87*(1), B35-B45. doi:10.1016/S0010-0277(02)00187-7

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733.

Patel, A. D., Iversen, J. R., & Hagoort, P. (2004). Musical syntactic processing in Brocas aphasia: A preliminary study. *Proceedings of the 8th International Conference on Music Perception and Cognition*.

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America, 124*, 3034–3047. doi:10.1121/1.2179657

Patel, A. D., & Morgan, E. (2016). Exploring cognitive relations between prediction in language and music. *Cognitive Science*, *41*(S2). 303-320. doi:10.1111/cogs.12411

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*. doi:10.3389/fpsyg.2012.00320

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*(7), 688-691.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117*(6).

Peretz, I., Vuvan, D., Lagrois, M., & Armony, J. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1664). doi:10.1098/rstb.2014.0090

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin & Review, 20*(2), 310-317. doi:10.3758/s13423-012-0344-5

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

Chapter 1: Introduction

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology, 28*, 142-149. doi:10.1016/j.conb.2014.07.005

Rebuschat, P., Rohrmeier, M., Hawkins, J. A., & Cross, I. (Eds.). (2012). *Language and music as cognitive systems*. Oxford: Oxford University Press.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *The Journal of Neuroscience, 31*(10), 3843-3852. doi:10.1523/jneurosci.4515-10.2011

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music, 5*(1), 35-53. doi:10.1080/17459737.2011.573676

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906-914. doi:10.1002/wcs.78

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*(4), 110-114. doi:10.1111/1467-8721.01243

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H. J., . . . Schulze-Bonhage, A. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *NeuroImage, 64*, 134-146. doi:10.1016/j.neuroimage.2012.09.035

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition, 58*(1), 75-125. doi:10.1016/0010-0277(95)00665-6

Scherer, K. R. (2013). Emotion in action, interaction, music, and speech. In M. Arbib (Ed.), *Language, music, and the brain: A mysterious relationship*. Cambridge, MA: MIT Press.

Schlaug, G., Marchina, S., & Norton, A. (2008). From singing to speaking: Why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music Perception, 25*(4), 315-323. doi:10.1525/MP.2008.25.4.315

Slevc, L. R., Faroqi-Shah, Y., Saxena, S., & Okada, B. M. (2016). Preserved processing of musical structure in a person with agrammatic aphasia. *Neurocase, 22*(6), 505-511. doi:10.1080/13554794.2016.1177090

Slevc, L. R., & Okada, B. M. (2014). Processing structure in language and music: A case for shared reliance on cognitive control. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-014-0712-4

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review, 16*(2), 374-381. doi:10.3758/16.2.374

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18*(5), 1169-1178. doi:10.1093/cercor/bhm149

Sun, Y., Lu, X., Ho, H. T., Johnson, B., & Thompson, W. F. (2015). Exploring music-syntactic processing and language-syntactic processing in congenital amusia using MEG and EEG. In J. Ginsborg, A. Lamont, M. Phillips, & S. Bramley (Eds.), *Proceedings of the 9th triennial conference of the European society for* *the cognitive sciences of music (ESCOM)* (pp. 765-770). Manchester, UK: Royal Nothern College of Music.

Sun, Y., Lu, X., Ho, H. T., & Thompson, W. F. (2017). Pitch discrimination associated with phonological awareness: Evidence from congenital amusia. *Scientific Reports, 7*.

Tallerman, M. (2015). *Understanding syntax*. New York: Routledge.

Thompson, W. F. (1996). A review and empirical assessment. *Journal of the American Musicological Society, 49*(1), 127-145. doi:10.2307/831956

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences, 109*(46), 19027-19032. doi:10.1073/pnas.1210344109

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*(1), 46-64. doi:10.1037/1528-3542.4.1.46

Tillmann, B., & Bigand, E. (2015). A commentary on "A commentary on: 'Neural overlap in processing music and speech'". *Frontiers in Human Neuroscience, 9*. doi:10.3389/fnhum.2015.00491

Trainor, L. J., & Trehub, S. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 18*(2), 394-402.

Trehub, S. (2016). Infant musicality. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed.). Oxford: Oxford University Press.

Trehub, S., & Trainor, L. J. (1993). Music and speech processing in the first year of life. *Advances in Child Development and Behavior, 24*.

Tzortzis, C., Goldblum, M., Dang, M., Forette, F., & Boller, F. (2000). Absence of amusia and preserved naming of musical instruments in an aphasic composer. *Cortex, 36*(2), 227-242. doi:10.1016/s0010-9452(08)70526-4

Van de Cavey, J., Severens, E., & Hartsuiker, R. J. (2017). Shared structuring resources across domains: Double task effects from linguistic processing on the structural integration of pitch sequences. *The Quarterly Journal of Experimental Psychology, 70*(8), 1633-1645. doi:10.1080/17470218.2016.1195852

Vos, S. H., Gunter, T. C., Kolk, H. H., & Mulder, G. (2001) Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology, 38*(1), 41-63.

Wilson, S., Parsons, K., & Reutens, D. (2006). Preserved singing in aphasia: A case study of the efficacy of melodic intonation therapy. *Music Perception, 24*(1), 23-36.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420-422. doi:10.1038/nn1872

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6*(1), 37-46. doi:10.1016/S1364-6613(00)01816

Chapter 2

Music and Language: Syntactic Interference Without Syntactic Violations

As reviewed in Chapter 1, the majority of studies investigating syntactic interference between music and language are based on stimuli containing syntactic violations. In Chapter 2, I therefore investigated Question 1—can syntactic interference be observed without violations of syntactic structure? Chapter 2 presents three experiments. Experiments 1 and 3 involved the simultaneous presentation of music and language. The primary task involved reading and then recalling word-lists and sentences. In Experiment 1, the secondary stimulus was background auditory stimuli. In Experiment 3, the secondary task involved detection of targets—an out-of-key note in melodic stimuli, and a "gong" sound in environmental stimuli. The results of Experiment 1 also had implications for Question 2—the connection between auditory streaming and syntactic processing. Thus, Experiment 2 in this chapter further examined the role of auditory streaming in syntactic processing.

This manuscript was co-authored by myself, Genevieve McArthur, and Bill Thompson. I contributed approximately 85% of the total work, including experimental design, stimuli creation, experiment programming, data collection, statistical analysis, and the preparation of the first draft of the manuscript. Bill Thompson and Genevieve McArthur contributed to experimental design, gave critical feedback throughout the process, helped with interpretation of data and provided helpful comments on the manuscript.

Chapter 2: Interference Without Violations

This chapter has been submitted as:

Fiveash, A., McArthur, G., & Thompson, W. F. (under review). Music and language: Syntactic interference without syntactic violations.

Music and Language: Syntactic Interference without Syntactic Violations

Anna Fiveash[1,3], Genevieve McArthur[2,3], & William Forde Thompson[1,3]

[1]Department of Psychology, Macquarie University

[2]Department of Cognitive Science, Macquarie University

[3]ARC Centre of Excellence in Cognition and its Disorders, Macquarie University

Corresponding author: Anna Fiveash, Department of Psychology, Macquarie University, NSW, 2109, Australia.

Email: anna.fiveash@mq.edu.au

Chapter 2: Interference Without Violations

Abstract

Music and language are complex hierarchical systems in which individual elements (e.g., chords, words) are systematically combined to form larger structural units. This property is known as syntactic structure. It has been suggested that music and language draw upon a pool of shared resources dedicated to syntactic processing. Most research supporting this suggestion has examined whether syntactic *violations* in one domain (out-of-key notes for music, grammatical errors for language) interfere with syntactic processing in the other domain. However, the distracting nature of syntactic violations may affect auditory processing in non-syntactic ways, accounting for reported interference effects. This study investigated whether syntactic structure in music *without* syntactic violations interferes with the processing of language syntax. In Experiment 1, we found syntactic interference in recall for sentences when participants listened to one-timbre melodies, but a reduced interference effect when listening to those same melodies with three alternating timbres. It was hypothesised that the alternating instruments interrupted auditory streaming, resulting in a less coherent syntactic representation, and therefore less interference with sentence recall. Experiment 2 tested this hypothesis. It found that participants were better able to discriminate two one-timbre melodies than two alternating timbre melodies. In Experiment 3, attention directed towards melodies compared to environmental sounds resulted in increased errors for sentence recall. This finding confirmed that attention plays a role in syntactic processing. Taken together, the results of these experiments suggest that syntactic processing resources, and not merely error processing mechanisms, are shared by music and language and are mediated by attention.

Keywords: music, language, syntax, attention, timbre, shared processing.

Chapter 2: Interference Without Violations

Public Significance Statement

The brain can process a large amount of information at the same time. However, there are processing limitations, especially when two types of information use similar cognitive resources. This research investigated the effect of background music on people's ability to read and remember language. We found that memory for what they read was reduced when music was played in the background, especially when people were paying attention to the music. We also found that music played with alternating instruments made it easier for people to remember written sentences than when music was played with just one instrument. This research has implications for the use of background music in situations where people read to retain information (such as working or studying), and sheds light on the way the brain processes multiple sources of information at the same time.

Chapter 2: Interference Without Violations

Music and language are two diverse, complex, rule-based systems that on the surface appear extremely different; however, current theory highlights a number of similarities between them. Despite having distinct functions and elementary units, music and language are both characterised by hierarchical structure in which discrete units (words, musical notes) are systematically combined to form larger structural units (sentences, musical phrases). The regularities governing how these elements are combined to form larger hierarchical sequences are collectively known as *syntax.* It has been shown that children learn syntactic rules in their native language quickly (Saffran, Senghas, & Trueswell, 2001), and similarly, children quickly develop an understanding of music syntax (Lamont, 2016). In addition, neuroimaging and behavioural research suggest similarities and overlap in syntactic processing between music and language (Koelsch, 2013; Patel, 2008), as well as transfer effects between them (Jentschke & Koelsch, 2009; Jentschke, Koelsch, Sallat, & Friederici, 2008).

There are two prominent theories that suggest music and language share cognitive resources for syntactic processing: the shared syntactic integration resource hypothesis (SSIRH; Patel, 2008), and the syntactic equivalence hypothesis (SEH; Koelsch, 2013). The SSIRH suggests that music and language draw upon domain-specific *representational networks* and domain-general *resource networks.* The representational networks are thought to hold information specific to music or language (e.g., verb categories in language, tonal knowledge in music). These networks can be selectively impaired, giving rise to double dissociations that are the hallmark of modularity arguments (Peretz & Coltheart, 2003). In contrast, domain-general resource networks are thought to facilitate the online syntactic integration of elements retrieved from representational networks. Syntactic integration is the process whereby incoming elements are combined with earlier elements to create a coherent sequence. The ease of integration is related to how expected the incoming elements are (Patel, 2008). If an element is unexpected in the syntactic context, more resources are required to integrate it into the sequence. These resource demands are reflected electrophysiologically by a brain response that occurs approximately 600 milliseconds (ms) after stimulus onset (the P600). The SSIRH is focused on syntactic integration, which is suggested to be a relatively "late" syntactic process (Patel, Gibson, Ratner, Besson, & Holcomb, 1998).

Similar to the SSIRH, the SEH suggests that music and language share syntactic processing resources. It further posits that any syntactically structured sequence (e.g., music, language, action, mathematics) shares resources for syntactic processing that are not shared by semantic processing or by acoustic deviance processing (Koelsch, 2013). In addition to integrational processes (labelled *structural reanalysis and revision*), Koelsch (2013) argues that "early" automatic processes of syntactic structure building are also shared between music and language (around 150ms post stimulus onset). Structural reanalysis and revision is the mechanism by which incoming elements are processed in relation to other elements in developing a syntactic representation. For a syntactic representation to develop, and for syntactic processing to occur, the incoming elements must first be perceived as one auditory stream (Bregman, 1990; Koelsch, 2011). Auditory streaming and early syntactic structure building are considered to be largely automatic, and can occur without attention (Bregman, 1990; Loui, Grent-'t-Jong, Torpey, & Woldorff, 2005; Maidhof & Koelsch, 2011).

In sum, both the SSIRH and the SEH suggest shared processing between music and language syntax, and predict interference effects (lowered performance) when music and language simultaneously place high demands on syntactic processing resources. Syntactic interference effects from music to language have been observed in both behavioural and neuroimaging studies. For example, behavioural studies have revealed that syntactic violations in sung sentences reduce comprehension for complex sentences (Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009), unexpected chords increase reaction times in a lexical decision task (Hoch, Poulin-Charronnat, & Tillmann, 2011), out-of-key chords increase reading times in garden path sentences (Slevc, Rosenberg, & Patel, 2009), and out-of-key chords reduce memory for complex sentences but not syntax-free word-lists (Fiveash & Pammer, 2014). Kunert, Willems, and Hagoort (2016) have also reported reverse interference effects from language to music: syntactic garden path sentences reduce harmonic closure judgements for final notes of a complex chord sequence compared to semantic garden path sentences and arithmetic calculations. Interestingly, music syntax processing does not appear to interfere with the processing of semantic anomalies in language (Hoch et al., 2011; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Slevc et al., 2009), and semantic anomalies elicit distinct brain responses compared to syntactic violations (Besson, Faita, Peretz, Bonnel, & Requin, 1998). These behavioural findings suggest that interference effects between music and language are specific to syntactic structure (but see Perruchet & Poulin-Charronnat, 2013).

The neuroimaging literature has also shown interference effects in the processing of music and language, supporting both the SSIRH and the SEH. Koelsch et al. (2005) found that the simultaneous presentation of syntactic violations in music and language produced a lowered left anterior negativity (LAN)—an event-related potential (ERP) component elicited with violations in language syntax. Koelsch et al. (2005) suggested that the decreased LAN reflected the cost of processing syntax in both domains simultaneously. Steinbeis and Koelsch (2008) found that the early right anterior negativity (ERAN), an ERP component elicited with errors in music syntax, was reduced in amplitude when paired with violations in language, suggesting that the processing of the language violation directly interfered with the processing of the music violation. A recent fMRI study presented participants with complex or simple sung sentences (object-extracted or subject-extracted relative clauses). Sentences were sung on melodies that contained a note that was in-key, out-of-key, or increased in loudness by 10 decibels (Kunert, Willems, Casasanto, Patel, & Hagoort, 2015). The authors subtracted brain activation to the simple sentences from brain activation to the complex sentences, and found overlap in Broca's area when the complex sentences were sung with an out-of-key note. This pattern did not occur with the loudness control, suggesting a syntax-specific interaction that occurred with integrational costs.

It is notable that the large majority of evidence for syntactic interference in music and language processing revolve around syntactic *violations*—out-of-key chords or notes in music, and grammatical errors in language. Such violation paradigms are problematic because interference could be explained by shared error detection mechanisms rather than shared syntactic processing resources per se. Although Kunert et al. (2015) argue that their out-of-key notes are not "errors" in a categorical sense, they are unexpected within the context and hence may trigger error detection mechanisms. To control for the possibility that syntactic violations are merely distracting, previous studies have included a control condition that does not involve a syntactic violation. Instead, a comparable source of distraction is introduced such as a change in timbre (Fiveash & Pammer, 2014; Slevc et al., 2009) or loudness (Fedorenko et al., 2009; Kunert et al., 2015). Such experiments have revealed interference in language processing when language is concurrently presented with an out-of-key element in music. This interference does not occur with the distracting control, suggesting that the interference effect is syntax-specific. However, Tillmann and

Bigand (2015) suggested that out-of-key notes or chords might produce greater distraction than control stimuli because they violate sensory *and* tonal expectations.

Thus, studies to date have not yet constructed a control condition that precisely matches the level of distraction associated with an out-of-key note or chord. This raises the possibility that syntactic violations in music and language lead to interference effects not because of shared capacity-limited syntactic processing resources, but because syntactic violations are highly distracting and engage error-detection mechanisms. This alternative interpretation indicates the need to evaluate shared syntactic resource models without introducing syntactic violations. If syntactic interference is still observed in the absence of syntactic violations in either domain, then it would seem more likely that syntactic processing, rather than distraction and error detection mechanisms, are shared between music and language. However, if syntactic interference between music and language is not observed in the absence of syntactic violations, it would seem more likely that stimulus violations are responsible for interference effects between music and language.

Experiment 1

Experiment 1 investigated whether syntactic interference effects between music and language are observed in the absence of syntactic violations. To test this hypothesis, four music conditions were created. The first condition consisted of *normal* melodies with syntactic structures that conformed to expectations (with syntax, low distraction). The second condition used *scrambled* melodies that disrupted global syntactic structure and hence did not conform to expectations (disrupted syntax, high distraction). Pilot testing showed that these stimuli were distracting, so a third (control) condition was created where each note in the normal melodies was randomly played on one of three different instruments to increase the level of distraction (with syntax, high distraction). This *timbre* melody condition was based on the changing-state hypothesis (Jones, Hughes, & Macken, 2010), which suggests that stimuli that is constantly "changing-state" (instantiated as a continual timbre change in this study) is considered to be more distracting than a "steady-state" sound. The fourth condition, *environmental sounds*, consisted of background soundscapes (e.g., ocean, jungle) that had no syntax or salient changing elements, and hence were considered to be minimally distracting (no syntax, low distraction).

These conditions were used to examine and tease apart the effects of music syntax and musical distraction on the ability to recall spoken language stimuli with syntax (sentences) and without syntax (word-lists). We decided to measure language recall because it is a sensitive measure of syntactic processing. To correctly recall a sentence, participants must parse the sentence, analyse the thematic relations (e.g., the order of events), and keep the general structure in mind. Considering the length of the sentences, participants would have to process them syntactically in order to recall all the words. For the shorter word-lists, participants are likely to use serial recall alone (and not syntactic processing), as syntactic relationships among the words are absent.

We predicted that (1) music containing syntax (normal and timbre melody conditions) would be associated with poorer recall for sentences (but not word-lists) than music with disrupted syntax or without syntax (scrambled melody and environmental sound conditions), (2) music that was distracting (scrambled and timbre melody conditions) would be associated with poorer recall for both sentences and word-lists compared to the normal melody and environmental sound conditions, and (3) music that contained syntax and was distracting (timbre melody condition) would be associated with the poorest recall for sentences. We also predicted that music with disrupted syntax that was also distracting (scrambled melody condition) would have no *syntactic* interference effects on recall of sentences, only distraction effects. However, this prediction was made with some caution since the continuous and unpredictable nature of this condition might encourage the cognitive system to try and discover or impose some level of syntax to make some sense of it. It is therefore possible that the scrambled condition may have had both syntactic interference and distraction effects on sentence recall, leading to the poorest performance in sentence recall.

Method

Participants. Fifty-four native-English speakers from Macquarie University participated in this study for course credit. Four participants were excluded: two due to recording error, and two due to error rates more than three standard deviations above the mean. This left 50 participants (M_{age} = 22.5, range: 18-68, 42 females). Participants had an average of 4.39 years of private music lessons (range: 0-28 years). Eighteen participants had five or more years of private lessons. Twelve participants considered themselves musical, 24 not musical, and 14 somewhat musical. Only 11 participants reported that they were currently musically active. All participants reported listening to music, with an average listening time of 112 minutes per day (*SD* = 55.72 mins).

Sample size was calculated using a G*Power analysis (Faul, Erdfelder, Lang, & Buchner, 2007) that considers effect sizes obtained in published research involving comparable conditions. Fiveash and Pammer (2014) found a small effect of background stimulus type on recall across both word-lists and sentences in a repeated measures ANOVA (η^2 = .05). Based on this effect size, and an α of 0.05, a sample size of 43 participants was required to achieve power of .95 (calculated using G*Power; Faul et al., 2007). We therefore aimed to recruit approximately 50 participants for our sample. This sample size is within the range used in previous syntactic interference studies (e.g., 32 in Hoch et al., 2011; 96 in Slevc et al., 2009). All participants were tested before data was scored, eliminating any chance of “optional stopping” (Rouder, 2014; Simmons, Nelson, & Simonsohn, 2011).

Design. The experiment was a 2 (language: sentences, word-lists) by 4 (music: normal, scrambled, timbre, environmental) within-subjects design. Following the main analysis, a subsequent analysis was conducted with musical training as a between-subjects variable (untrained = zero years of private music lessons; trained ≥ 5 years private music lessons). Five or more years of private music lessons was considered “musically trained” based on previous literature (e.g., see Krumhansl & Kessler, 1982; Singer et al., 2016; Vuvan, Prince, & Schmuckler, 2011). Although musical training is a continuous variable, this categorisation allowed for a clear distinction between a group of participants with no musical training and one with substantial musical training. It should be noted that this secondary analysis only included a subset of participants (n = 39), as participants with one to four years of musical training were excluded.

Language stimuli consisted of 60 complex sentences and 60 word-lists. Music stimuli consisted of 30 stimuli in each of the four conditions (i.e., 120 in total). Music and language stimuli were randomised for each participant so that the 30 stimuli in each music condition were randomly paired with 15 word-lists and 15 sentences. The timing of stimuli presentation for each trial is shown in Figure 1. There were 120 trials in total. Stimuli were presented via Matlab (version 2014b; Mathworks) using Psychtoolbox (version 3.0.12; Brainard, 1997).

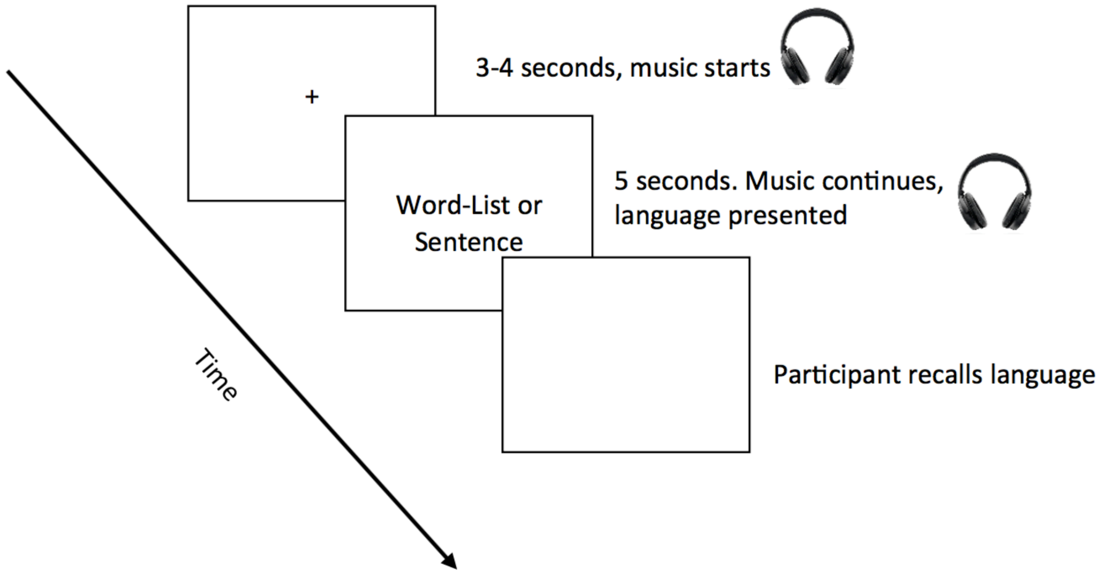

Figure 1. Timing of stimulus presentation. Headphones represent auditory stimulus presentation.

Music stimuli. The *normal melody* condition (with syntax, low distraction) consisted of 30 single line musical instrument digital interface (MIDI) melodies. All stimuli were approximately 9 seconds long, 120 beats per minute (bpm), were composed in the keys of C, G, D, and A major, and had an average of 23.9 notes (range: 18-30). Melodies were played through MIDI instrument Steinway grand piano and exported using GarageBand. See Figure 2a.

The *scrambled melody condition* (disrupted syntax, high distraction) was created by taking each note and its duration from each normal melody and randomising note order. Each melody in the normal melody condition therefore had a scrambled version with the same notes and note durations but with disrupted global syntax (both melodic and rhythmic syntax were disrupted). Scrambled music has previously been used as a “non-syntactic” comparison condition, as it holds constant the total acoustic information available, but disrupts syntactic structure (Abrams, Bhatara, Ryali, Balaban, Levitin, & Menon, 2011; Levitin & Menon, 2003). In previous work, however, scrambled stimuli have been created by splicing recordings into 250-350ms chunks of sound, regardless of whether the segmentation interrupted individual notes or chords, and then re-splicing these segments in a random order. Thus, the procedure not only scrambled syntactic structure, but also disrupted the processing of individual tones and chords every 250-350ms, undermining fundamental mechanisms of auditory streaming. As such, observed differences in responses to fully syntactic stimuli and re-spliced scrambled stimuli could reflect differences in syntactic processing, differences in the engagement of auditory streaming, or both. In contrast, we employed randomised stimuli that disrupted syntactic structure, but retained all of the discrete elements contained within the original melodies (see Figure 2b).

The *timbre melody condition* (with syntax, high distraction) was included to match the level of distraction of the scrambled melody condition while maintaining syntactic structure. To achieve this, the notes in the normal melodies were played through three different MIDI instruments (Steinway grand piano, acoustic guitar, and vibraphone). An external random number generator was used to determine which of the three instruments would play each note, and no instrument played more than three notes in a row. See Figure 2c.

The *environmental sound condition* (no syntax, low distraction) was included as a control. Environmental stimuli were downloaded from the www.sounddogs.com website. Thirty sounds were chosen as background ambient or common sounds (e.g., jungle background, train station ambience, ocean noises etc.). Stimuli were shortened and normalised for loudness in respect to the other stimuli, and there were no salient distracting features.

a)

b)

c)

Figure 2. a) normal melody, b) scrambled melody, c) timbre melody (where black = piano, green = guitar, red = vibraphone). Not pictured: environmental sounds.

Language stimuli. There were two types of language stimuli: complex sentences and word-lists. Object-extracted sentences (10-16 words, 50-68 characters) were adapted from Fedorenko, Gibson, and Rohde (2006) and Fedorenko et al. (2009). Five-word word-lists (30-33 characters, 8-10 syllables) were created from randomisations of the sentences' content words. Word-lists were created this way to try and maintain a similar level of difficulty across the sentences and word-lists. As there were not quite enough content words to finish the 60 word-lists, extra words were taken from the word-lists in Fiveash and Pammer (2014).

Procedure. Participants first read and signed the information and consent form, and then completed a brief musical education and preference questionnaire. Participants were told they would hear four different types of music stimuli that would be paired with either a sentence or a word-list, and that their task was to recall the sentence or word-list out loud once it disappeared from the screen. Participants were given practice trials where they heard examples of the four music conditions, and read and recalled sentences and word-lists. After the practice trials, the experimenter confirmed verbally and through participant responses that the participant understood the task, and the experiment proper began. All recordings were transcribed for scoring purposes.

Scoring. Sentences and word-lists were scored in a similar way to Mann, Liberman, and Shankweiler (1980), where the final score was the sum of all omissions, substitutions, variants, additions, and reversals (see Table 1). Each type of error was evenly weighted in the error score calculation, receiving a score of 1. Higher scores reflect more errors, and therefore lower recall of the language stimuli. All stimuli were blind scored so that markers were unaware of condition. One marker scored all responses and a second marker scored 31 of the 50 responses. The markers agreed on 94.4% of responses. Any discrepancies were discussed and a mutual agreement was reached.

Analysis. Given that sentences comprised more words than word-lists, sentences were associated with higher error scores (omissions and inaccurate recall) than word-lists. For each type of language stimuli (sentences and word-lists), we therefore conducted two analyses of variance (ANOVAs) to analyse sentences and word-lists separately, as they were not directly comparable. In each ANOVA, we used a criterion (i.e., *p* value) of 0.05, and any statistically significant main effects or interactions were further explored using Holm-Bonferroni corrected pairwise comparisons. Cohen's *d* effect sizes for these comparisons are reported based on repeated measures data, taking into account correlations between conditions.

The primary analysis consisted of repeated measures ANOVAs that compared the effects of the four music conditions (normal, scrambled, timbre, and environmental) on language stimuli separately (either sentences or word-lists). We expected an additive effect of syntax and distraction in recall for sentences, and only distraction effects to be observed in recall for word-lists.

To explore whether musical training had an impact on sentence recall, we ran an ANOVA on a subset of participants (n = 39), with the same factor of music condition (normal, scrambled, timbre, and environmental), and the added between-subjects factor of musical training (untrained, n = 21; trained, n = 18). Statistically significant main effects or interactions were further explored using pairwise comparisons (Holm-Bonferroni corrected).

Table 1

Scoring of Sentences and Word-lists.

Error Type	Description
Omissions	No recall.
Substitutions	Word substituted for another. E.g., synonyms, or close in pronunciation.
Variants	Same word, just different variants of it. Also used if one letter was different.
Additions	Words not in original stimuli.
Reversals	Word in wrong position. If following words were in the same *relative* order after the reversal, these were scored as correct.

Results

Sentences. Error scores for sentence recall across the four music conditions are presented in Figure 3. The main effect of music condition was significant, $F(3, 147) = 47.00$, $p < .001$, $\eta^2 = .49$. There were significant differences between each condition, as confirmed by pairwise comparisons (Holm-Bonferroni corrected *p* values reported for six comparisons). The scrambled melody condition (disrupted syntax, high distraction; $M = 2.49$, $SD = 1.16$) was associated with lower sentence recall than the normal melody condition (with syntax, low distraction; $M = 2.13$, $SD = 1.25$), $t(49) = 3.57$, $p' = .002$, $d = .51$, the timbre melody condition (with syntax, high distraction; $M = 1.76$, $SD = 1.20$), $t(49) = 7.00$, $p' < .001$, $d = 1.0$ and the environmental sound condition (no syntax, low distraction; $M = 1.19$, $SD = .75$), $t(49) = 10.20$, $p' < .001$, $d = 1.58$.

Looking beyond the scrambled condition, an interesting trend was observed. As predicted, the environmental sound condition (no syntax, low distraction) was associated with significantly fewer errors than both the normal melody condition (with syntax, low distraction), $t(49) = 6.83$, $p' < .001$, $d = 1.09$, and the timbre melody condition (with syntax, high distraction), $t(49) = 5.12$, $p' < .001$, $d = .86$. However, the normal melody condition was associated with significantly more errors than the timbre melody condition, $t(49) = 3.70$, $p' = .002$, $d = .52$. This finding was unexpected because the timbre melody condition was hypothesised to produce additive effects of syntactic interference and distraction. This unexpected result will be discussed further in the Discussion below.

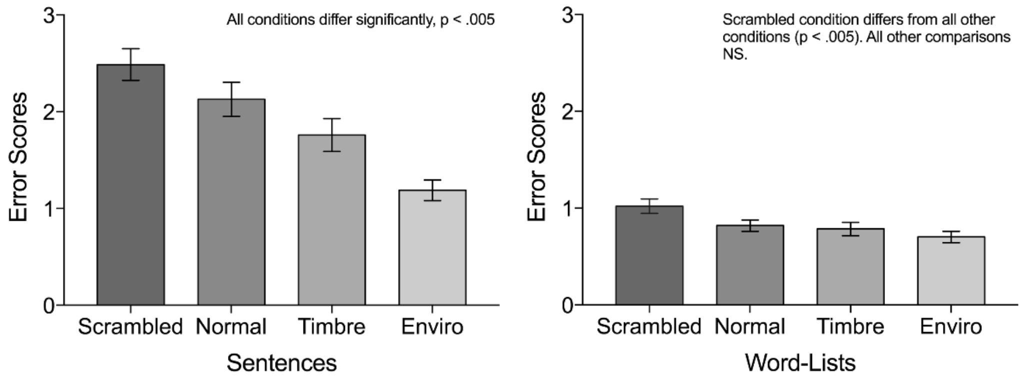

Figure 3. Error scores in sentence and word-list recall across the four music conditions. Enviro refers to the environmental condition. Error bars indicate one standard error either side of the mean. NS = non-significant.

In the secondary analysis, the main effect of music condition was significant $F(3, 111) = 39.09$, $p < .001$, $\eta^2 = .51$, and there was a condition by musical training interaction, $F(3, 111) = 2.99$, $p = .034$, $\eta^2 = .08$. As predicted, there was also a significant between-subjects effect of musical training, $F(1, 37) = 6.67$, $p = .014$, $\eta^2 = .15$. To explore the interaction and between-subjects effects, we ran independent-samples *t*-tests with musical training as the grouping factor. Holm-Bonferroni adjusted *p* values are reported for four comparisons. These comparisons showed that musically trained participants made significantly fewer errors in sentence recall when listening to the normal melody (untrained: $M = 2.69$, $SD = 1.24$; trained: $M = 1.59$, $SD = .91$), $t(37) = 3.1$, $p' = .016$, $d = 1.01$, and scrambled melody (untrained: $M = 2.86$, $SD = 1.18$; trained: $M = 2.04$, $SD = .77$), $t(37) = 2.54$, $p' = .045$, $d = .82$ conditions. There were no differences between groups in the timbre melody (untrained: $M = 2.12$, $SD = 1.45$; trained: $M = 1.41$, $SD = .81$), $t(32.13) = 1.94$, $p' = .12$[1], or the environmental sound conditions (untrained: $M = 1.3$, $SD = .82$; trained: $M = .97$, $SD = .56$), $t(37) = 1.52$, $p' =$

[1] Levene's test indicated unequal variances ($F = 11.46$, $p = .002$), so degrees of freedom were adjusted from 37 to 32.13.

.14. These results show that for the normal and scrambled conditions, musically trained participants made significantly fewer errors in sentence recall than participants without musical training. The musically trained group also had fewer errors than the untrained group for the timbre and environmental conditions, though these differences were not significant.

Word-lists. A repeated measures ANOVA was conducted to examine error scores for word-list recall across the four music conditions. The main effect of music condition was significant, $F(3, 147) = 11.53$, $p < .001$, $\eta^2 = .19$. Figure 3 and pairwise comparisons (Holm-Bonferroni corrected p values reported for six comparisons) revealed that this significant result was driven by the scrambled melody condition (M = 1.02, SD = .53), which resulted in significantly more errors than the other three conditions: normal melody (M = .82, SD = .42), $t(49) = 3.63$, $p' = .004$, $d = .52$, timbre melody (M = .79, SD = .49), $t(49) = 3.99$, $p' < .001$, $d = .55$, and environmental sounds (M = .70, SD = .41), $t(49) = 5.51$, $p' < .001$, $d = .81$. The other differences between conditions were not significant: normal and timbre melodies, $t(49) = .61$, $p' = .55$, normal melodies and environmental sounds, $t(49) = 2.34$, $p' = .07$, timbre melodies and environmental sounds, $t(49) = 1.42$, $p' = .32$, suggesting that recall was similar across these three conditions for the word-lists. This finding differs from the sentence results, and suggests that syntactic interference occurs for sentences but not for word-lists. An exploratory analysis with musical training as a between-subjects factor still showed a main effect of condition, $F(3, 111) = 10.12$, $p < .001$, $\eta^2 = .22$, no interaction, $F(3, 111) = .28$, $p = .84$, and no between-subjects effect of musical training on word-list recall, $F(1, 37) = 3.91$, $p = .056$.

The scrambled condition. We were interested to find that the scrambled condition resulted in the poorest recall for both sentences and word-lists. This result could be interpreted in two ways. First, the unexpected nature of the randomised notes may have been too distracting to perform the language task and hence led to poorer performance compared to the other conditions. This interpretation would explain why the scrambled condition resulted in significantly poorer performance for both sentence and word-list recall. To examine this possibility further, we used word-list error scores as a baseline, and calculated the difference score of sentences minus word-lists for each participant. Any differences between conditions should be a result of syntactic interference rather than distraction effects. An ANOVA, $F(3, 147) = 23.69$, $p < .001$, $\eta^2 = .33$, indicated that the main effect of music condition was significant. Pairwise comparisons (Holm-Bonferroni corrected *p* values reported for six comparisons) showed no significant difference between the normal ($M = 1.31$, $SD = 1.11$) and scrambled melody ($M = 1.47$, $SD = .96$) conditions, $t(49) = 1.31$, $p' = .20$, suggesting that the main difference between these two conditions may be in the higher level of distraction for the scrambled melody stimuli. There was also a significant difference between the normal and timbre melody ($M = .97$, $SD = .91$) conditions, $t(49) = 2.92$, $p' = .01$, $d = .43$, suggesting that the difference in timbre between the sentences and word-lists was an effect of syntax, not merely a lower level of distraction. All other comparisons remained significantly different: normal melody and environmental sounds ($M = .49$, $SD = .64$), $t(49) = 5.39$, $p' < .001$, $d = .81$, scrambled and timbre melodies, $t(49) = 4.65$, $p' < .001$, $d = .66$, scrambled melody and environmental sounds, $t(49) = 7.13$, $p' < .001$, $d = 1.04$, and timbre melody and environmental sounds, $t(49) = 4.11$, $p' < .001$, $d = .60$.

A second explanation for why the scrambled condition led to the poorest recall is that participants may have tried to impose a syntactic structure on the sequence, therefore engaging syntactic processing resources to a greater extent. We did not intend for the scrambled condition to contain syntactic “violations”; however, it is possible that by disrupting global syntax, we also violated temporal and melodic expectations. By definition, violations of expectancies vary with the complexity of melodies. To confirm that the scrambled sequences were more complex than normal sequences, we used the MIDI toolbox (Eerola & Toiviainen, 2004) to calculate melodic complexity for the normal and scrambled melodies, based on pitch and rhythmic aspects of the MIDI files (Eerola, 2016). According to this model, and confirmed by a paired-samples *t*-test, the scrambled melodies ($M_{\text{complexity}} = 4.19$, $SD = 1.88$) were significantly more complex than the normal melodies ($M_{\text{complexity}} = 3.83$, $SD = .22$), $t(29) = 30.12$, $p < .001$, $d = 1.74$. This difference in complexity suggests that another reason scrambled melodies resulted in the poorest sentence recall is because of increased use of syntactic processing resources to try and make sense of the unexpected notes. Considering the difference scores and melodic complexity ratings together, it is likely that the scrambled condition both increased distraction *and* increased syntactic processing effects.

Discussion

Experiment 1 investigated the effects of four different music conditions (normal, scrambled, timbre, and environmental) on recall for complex sentences (with syntax) and word-lists (no syntax). As predicted, participants made significantly more errors in sentence recall when the sentences were paired with normal melodies (with syntax, low distraction) or timbre melodies (with syntax, high distraction) than when they were paired with environmental sounds (no syntax, low distraction). These effects were not evident for the word-lists. These outcomes suggest that syntactic interference can occur between music and language processing even without syntactic violations. This effect does not appear to depend on musical training, though participants with five or more years of musical training made fewer errors in recalling sentences than participants with no musical training in the normal and scrambled melody conditions. There were no differences between musically trained and untrained participants in word-list recall.

The results of Experiment 1 provided us with two further insights. First, the scrambled condition (disrupted syntax, high distraction) led to the poorest recall across both sentences and word-lists. An analysis on the difference between sentence and word-list error scores suggested that the scrambled condition was more distracting than the other conditions. A complexity analysis revealed that the scrambled melody condition was also more melodically complex than the normal melody condition. This difference in complexity suggested that the scrambled condition could have engaged syntactic resources to a greater extent as participants tried to make sense of the more complex and unexpected incoming information. We suggest a combination of these two factors that should be teased apart in future research. Furthermore, it should be noted that the scrambled condition also disrupted rhythmic expectations, which are an important aspect of syntactic processing (Lerdahl & Jackendoff, 1983). While our main focus was on melodic syntax, the disruption of rhythmic syntax could have also affected the current results.

Second, Experiment 1 revealed that although the timbre melody condition was associated with significantly more errors than the environmental condition (as expected), it unexpectedly led to significantly *fewer* errors than the normal condition. As this effect did not occur for word-lists, it appears that the alternating timbres changed the way participants processed the syntax in the music, and reduced the interference effect when paired with language. This result might have occurred because the changing timbres disrupted auditory streaming processes (Bregman, 1990), resulting in a less coherent syntactic representation of the melody. This disrupted representation may have freed up resources for syntactic processing of sentences. It is also possible that the alternating timbres resulted in attention being drawn *to* the timbres, and *away* from the syntax, again, freeing up resources for syntactic processing of sentences. These possibilities lead to the prediction that alternating timbres within a melody results in poorer syntactic processing. This hypothesis was tested in Experiment 2.

Experiment 2

It has been hypothesised that stimuli that are perceptually grouped as originating from different sources, or stimuli that change state (e.g., change acoustically in some way) require more attention to process (changing-state hypothesis; Hughes, Hurlstone, Marsh, Vachon, & Jones, 2013). It has also been suggested, in both the changing-state hypothesis, and with Gestalt grouping principles, that incoming auditory information is grouped via auditory streaming processes into meaningful units, and that this process has a cognitive processing cost (Bregman, 1990; Deutsch, 2013; Zeamer & Fox Tree, 2013). When combined, these hypotheses suggest that the timbre condition in Experiment 1 may have both induced an increased cognitive cost, and led to an impaired syntactic representation of the melody, because of disrupted auditory streaming processes (Bregman, 1990). Given that a syntactic sequence can only emerge from a coherent auditory stream of musical elements, any disruption of auditory streaming should limit the level of syntactic processing that occurs. The weakened music syntax processing, in turn, should free up resources for processing *language* syntax, resulting in improved sentence recall.

To test this suggestion, a same-different paradigm was used in Experiment 2 to investigate whether music syntax is processed to a greater extent in one-timbre melodies (normal melody condition) compared to three-timbre melodies (timbre melody condition). If listeners are more sensitive to syntactic changes in normal melodies compared to timbre melodies, this result would suggest that changing timbres interrupt and reduce syntactic processing. This result would explain why the timbre condition was associated with fewer errors in sentence recall than the normal condition in Experiment 1.

Method

Participants. Forty-three participants were recruited from Macquarie University and participated for course credit. Two participant's data were lost due to technical errors, leaving 41 participants (M_{age} = 22.10 years, age range: 18-70, 30 females). Participants had an average of 4.5 years of private music lessons (range: 0-20 years). Fifteen participants had five or more years of private music lessons. Seven participants indicated they were musicians, 24 indicated they were non-musicians, and 10 indicated they were "somewhat" musicians. Fourteen participants reported that they were currently musically active. All participants reported listening to music, with an average listening time of 144 minutes per day (*SD* = 107.7 mins).

Previous same-different experiments included a range of sample sizes (e.g., sample sizes of 30, 14, 21, and 64 respectively; Croonen, 1994; Dowling, 1971, 1978; Weiss, Trehub, & Schellenberg, 2012), and there are no existing studies that compare melodies with alternating timbres in one auditory stream as in the current study. We therefore ran a G*Power analysis with a medium-to-large effect size, (d_z = .65), and an α of 0.05 which determined 33 participants were needed to achieve power of .95. Given our uncertainty about the size of the effect, we ran an additional 10 participants to ensure we had adequate power to detect an effect. Data collection was finalised before data was analysed, ensuring there was no optional stopping (Rouder, 2014; Simmons et al., 2011).

Design. Experiment 2 was a same-different task, with a 2 (melodies: same, different) by 2 (music condition: normal, timbre) within-subjects design. As in Experiment 1, a subsequent analysis was conducted on a subset of participants (n = 33) with musical training as a between-subjects variable (untrained = zero years of private music lessons; trained ≥ 5 years private musical training).

Stimuli. The melodies were the same as Experiment 1, which were played either with a piano (normal melody condition) or with three alternating instruments (timbre melody condition). In half the trials, the melodies were the same, and in half the trials they were different. In *different* trials, two melodies could differ by an *altered* note (i.e., a nearby in-key note) or by a *violation* note (i.e., a nearby out-of-key note). Note changes were always on the first or third beat (the strong beats), and in the second or third bars of the four bar melodies. The altered and violation manipulations were included to ensure optimal sensitivity to syntactic processing, as detecting an altered note that is in-key is considerably harder than detecting an altered note that is out-of-key. Including both manipulations therefore provided a range of difficulty in same-different judgements.

The pairings were created using Audacity software, and there was a 2-second break between melodies, as in the Montreal Battery of Evaluation of Amusia (MBEA; Peretz, Champod, & Hyde, 2003). Same-different pairs had six different possible combinations: normal-normal (same), altered-altered (same), violation-violation (same), normal-altered (different), normal-violation (different), altered-violation (different). For each of the pairings, melodies were presented in both possible orders (e.g., normal-violation and violation-normal). See Figure 4 for an example.

Figure 4. Examples of melodies in the same-different paradigm. (a) is an example of a violation-normal *different* trial. (b) is an example of a violation-violation *same* trial.

Procedure. Participants first read and signed the information and consent form, and then completed a brief musical education and preference questionnaire. Participants were instructed that they would hear two melodies—one after the other—which would either be both played by piano alone, or by alternating timbres. They were told that the timbres would not change, that there would only be a one-note difference (either in-key or out-of-key) between melodies, and that they should indicate whether there was a difference by pressing the same (z) or different (m) key on the keyboard. Practice trials contained examples of both the timbre and normal conditions, as well as examples of altered and violation melodies. After ensuring that the participant understood the task, the experiment proper began. The experiment consisted of 60 normal melody pairs and 60 timbre melody pairs (i.e., 120 trials in total) presented via Matlab (version 2014b, Mathworks) and Psychtoolbox (Brainard, 1997). All pairings were randomised so that order of presentation was different for each participant. Participants had a break after every 30 trials. The dependent variables were accuracy and reaction time (RT). The whole process took approximately 50 minutes.

Analysis. To calculate how sensitive participants were to differences in stimuli for the normal and timbre melody conditions, d prime (d') values were calculated using signal detection theory (Stanislaw & Todorov, 1999). D prime measures sensitivity to signal versus noise without response bias. A *hit* was recorded when the correct response was different, and the participant answered different. A false alarm was recorded when the correct response was same, and the participant answered different. Z scores were calculated, and z (false alarms) were subtracted from z (hits) to calculate the d' value for each participant in the normal and timbre melody conditions.

To assess whether there were differences in sensitivity (d' scores) and RT (milliseconds, ms) in same-difference judgements between the normal and timbre conditions, we ran paired-samples *t*-tests. To explore whether there was a difference depending on musical training (untrained, n = 18; trained, n = 15), we ran independent-samples *t*-tests with musical training as the grouping factor.

Results

Sensitivity analyses. A paired-samples *t*-test revealed that the d' values for the normal melody condition (M = .74, SD = .46) were significantly higher than the d' values for the timbre melody condition (M = .55, SD = .43), $t(40)$ = 3.08, p = .004, d = .48. This finding suggests that participants were more sensitive to differences between melodies in the normal melody condition compared to the timbre melody condition.

The independent-samples *t*-tests (Holm-Bonferroni adjusted *p* values reported for two comparisons) showed that same-different judgements in the normal melody condition were marginally higher for musically trained (M = .94, SD = .45) compared to untrained participants (M = .57, SD = .45), $t(31) = 2.32$, p' = .054. This difference was not significant when controlling for multiple comparisons. There was no difference for same-different judgements in the timbre melody condition depending on musical training: trained (M = .66, SD = .28), untrained (M = .47, SD = .45), $t(28.91) = 1.4$, p' = .17[2]. These results show that musical training did not influence same-different judgements when multiple comparisons were controlled for.

Reaction times. Trials that were more than 3 *SD* above the grand RT mean were excluded from average RT score calculations. A paired-samples *t*-test revealed that the mean RT (ms) for the normal melody condition (M = 694.7; SD = 266.81) was significantly shorter than the timbre melody condition (M = 754.7; SD = 265.20), $t(40) = 3.86$, $p < .001$, d = .60. The exploratory, independent-samples *t*-tests (Holm-Bonferroni adjusted *p* values reported for two comparisons) revealed no significant difference in RT in the normal melody condition for the musically trained (M = 615.42, SD = 317.48) compared to the untrained group (M = 679.99, SD = 222.79), $t(31)$ = .69, p' = .50. There was also no difference between the musically trained (M = 643.25, SD = 293.00) and untrained group (M = 753.52, SD = 206.25) in the timbre melody condition, $t(31) = 1.23$, p' = .43.

Discussion

Experiment 2 showed that differences in syntax between two melodies were detected more quickly and accurately when played with a single instrument compared

[2] Levene's test indicated unequal variances (F = 4.24, p = .048), so degrees of freedom were adjusted from 31 to 28.91.

to alternating instruments. Exploratory analyses that compared musically trained and untrained groups revealed no significant differences depending on musical training when multiple comparisons were controlled for. However, participants with five or more years of training were marginally more accurate at same-different judgements in the normal melody condition compared to participants with no musical training.

Melodies with alternating timbres were initially included in Experiment 1 to create a higher level of distraction in the musical stimuli, in line with the changing-state hypothesis (Hughes et al., 2013; Jones et al., 2010). However, the results of Experiments 1 and 2 combined suggest that alternating timbres may reduce syntactic processing in music, possibly through both a disruption to the auditory stream, and less attention directed to syntactic processing. This reduced syntactic processing may free up resources to process language (as observed in Experiment 1), and make it more difficult to compare syntactic structure between two melodies (as observed in Experiment 2). The importance of attention and auditory streaming in the shared syntactic processing of music and language was therefore explored more thoroughly in Experiment 3.

Experiment 3

Attention is an important element in processing incoming auditory information, and it appears to modulate the amount of music syntax processing that occurs. Maidhof and Koelsch (2011) found that although brain responses to syntactic violations in music can be elicited without attention, they are strongest when participants only hear music, and decrease when participants are also simultaneously attending to other auditory stimuli. Furthermore, Loui et al. (2005) found that the early anterior negativity (EAN) brain response to syntactic violations was elicited in both attended and unattended conditions, but was more pronounced in the attended condition. These findings suggest an important role of attention in syntactic processing.

Experiment 3 was designed to determine whether attention modulates the interaction between music and language syntax within an auditory streaming framework. To this end, melodies were presented to one ear and environmental sounds were presented to the other ear, enforcing streaming of incoming auditory information. Participants were instructed to direct their attention to one of these streams whilst reading word-lists, basic sentences, or complex sentences. We predicted that recall for complex sentences would decrease when participants were attending to melodies compared to environmental sounds due to syntactic interference effects when simultaneously processing music and language. We also expected some interference in recall for basic sentences when participants were attending to the melody; however, we expected this interference to be smaller than for the complex sentences, as basic sentences have less complex and demanding syntax.

Method

Participants. Seventy-four native-English speakers participated for either course credit or $15 from Macquarie University. Four people were excluded due to computer error, and one outlier was excluded due to poor recall (3 *SD* below the mean across all conditions). To ensure that participants were paying attention to the correct ear, a threshold of 50% target detection in both the melody and environmental sound streams was set. Due to this threshold (and an examination of d' values), 20 participants were excluded. This left 49 participants in total (M_{age} = 22.12, age range: 18-43, 36 females). Of these, participants had an average of 5.6 years of private music lessons (range: 0-22 years). Twenty-one participants had five or more years of private music lessons. Fourteen indicated that they were musicians, 25 indicated they were nonmusicians, and 10 indicated they were “somewhat” musicians. Twenty-two participants reported that they were currently musically active. All participants reported listening to music, with an average listening time of 104 minutes per day (*SD* = 65.9 mins).

Sample size was calculated as in Experiment 1, based on a G*Power analysis incorporating the effect size ($\eta^2 = .05$) from Fiveash and Pammer (2014), an α of 0.05, and a required power of .95. This power analysis suggested a sample size of 43. In line with Experiment 1, and because this experiment included different conditions to Fiveash and Pammer (2014), we aimed for a sample size of 50. All participants were tested before recall was scored, to avoid optional stopping (Rouder, 2014; Simmons et al., 2011). However, target detection was monitored throughout testing to ensure there were 50 participants who scored above the target detection cut-off.

Design. Experiment 3 was a 2 (attention: melodies, environmental sounds) by 3 (language: word-lists, basic sentences, complex sentences) within-subjects design. As in Experiments 1 and 2, a subsequent, exploratory analysis with a subset of participants (*n* = 35) was conducted with musical training as a between-subjects variable (untrained, trained).

In half the trials, participant’s attention was directed to the melody, and in the other half of the trials, their attention was directed to the environmental sounds. Attention was directed by asking participants to detect a target that was either an out-of-key note (in the melody) or a gong sound (in the environmental sounds). Targets were present in 20% of trials. Language recall on these trials was not analysed. Ear presentation (i.e., the ear in which the melody or environmental sounds were presented) was counterbalanced across participants, and stimuli order was randomised for each participant. There were 120 trials in total, and stimuli were presented using Matlab (version 2014b, Mathworks) and Psychtoolbox (Brainard, 1997).

Language stimuli. There were 40 word-lists, 40 basic sentences, and 40 complex sentences. Word-lists were the same as Experiment 1. Eighty sentences (10-14 words, 12-20 syllables) were adapted from previous work investigating subject-extracted (basic) versus object-extracted (complex) sentences (Fedorenko et al., 2006; Fedorenko et al., 2009; Staub, 2010). Sentences were then randomised so that 40 sentences maintained a basic structure (e.g., *The crook who warned the thief fled the town the next morning*), and 40 sentences were manipulated to have a complex structure (e.g., *The scout who the coach punched had a fight with the manager*). Complex sentences have a higher processing cost because the verb (e.g., *punched*) has to be held in memory for a longer time before it can be combined with the subject (e.g., *the scout*; Fiebach, Schlesewsky, & Friederici, 2001). Basic and complex sentences were included to examine whether there was an effect of syntactic complexity.

Melodic stimuli. The 30 melodies from Experiments 1 and 2 were used. The violation stimuli from Experiment 2 were used as target melodies. In these stimuli, the out-of-key notes were always in the last two bars, on a one or three beat. Twelve of the 60 melodies contained a target, and targets were spread evenly throughout the language conditions.

Environmental stimuli. Environmental sounds were taken from the NESSTI: norms for environmental sound stimuli collection (Hocking, Dzafic, Kazovsky, & Copland, 2013). This collection had 110 sounds from multiple categories. All sounds were 1 second long. Sounds that could interfere with melodic stimuli (e.g., human sounds, musical sounds), or that had a noticeably lower intensity and reduced sound spectrum compared to the other recordings were not used. This left 69 sounds across different environmental categories. All sounds were normalised so they were at a similar loudness level to the melodies. The 69 one-second environmental sound chunks were then randomised into groups of eight. This process formed 30 different environmental sound sequences, each containing eight individual environmental sounds (e.g., telephone, pinball machine, frog, bee, train, cow, ocean, crow). Individual environmental sounds could be used in multiple sequences but were not repeated within a single sequence. Twelve of the environmental sound sequences contained the target gong sound. The gong always occurred in one of the last four positions.

These 30 different environmental sound sequences were then randomly paired with the 30 melodies four times to create 120 stimuli combinations. Each melody was repeated four times, and each sequence of eight environmental sounds was repeated four times, but the same pairing was never repeated. The same melody or set of environmental sounds was never presented in succession, and no stimuli ever contained targets in both stimuli at the same time. Two sets of stimuli were created, one with the melody in the left ear, and one with the melody in the right ear.

Procedure. Participants read and signed the information and consent form, and completed a brief musical education and preference questionnaire. The procedure was explained, examples of the gong and out-of-key notes were played, and participants practiced the task. When it was clear that they understood the task, the experiment proper began. On each trial, the participant's attention was directed to one auditory stream by being asked either “is there a gong?” or “is there an out-of-key note?” Direction of attention (to melodies or to environmental sounds) was randomised throughout trials. Each melody initially played for 2-3 seconds (depending on whether its duration was 8 or 9 seconds) and continued while a word-list, basic sentence, or complex sentence appeared on the screen for 6 seconds (note: we increased the language stimuli presentation by 1 second for Experiment 3 because the auditory task was more demanding than Experiment 1). At the end of each trial, the participant recalled the language stimulus out loud, and their response was recorded. Participants were then asked to indicate whether they detected the target. The procedure is shown in Figure 5. The experiment took approximately 50 minutes in total.

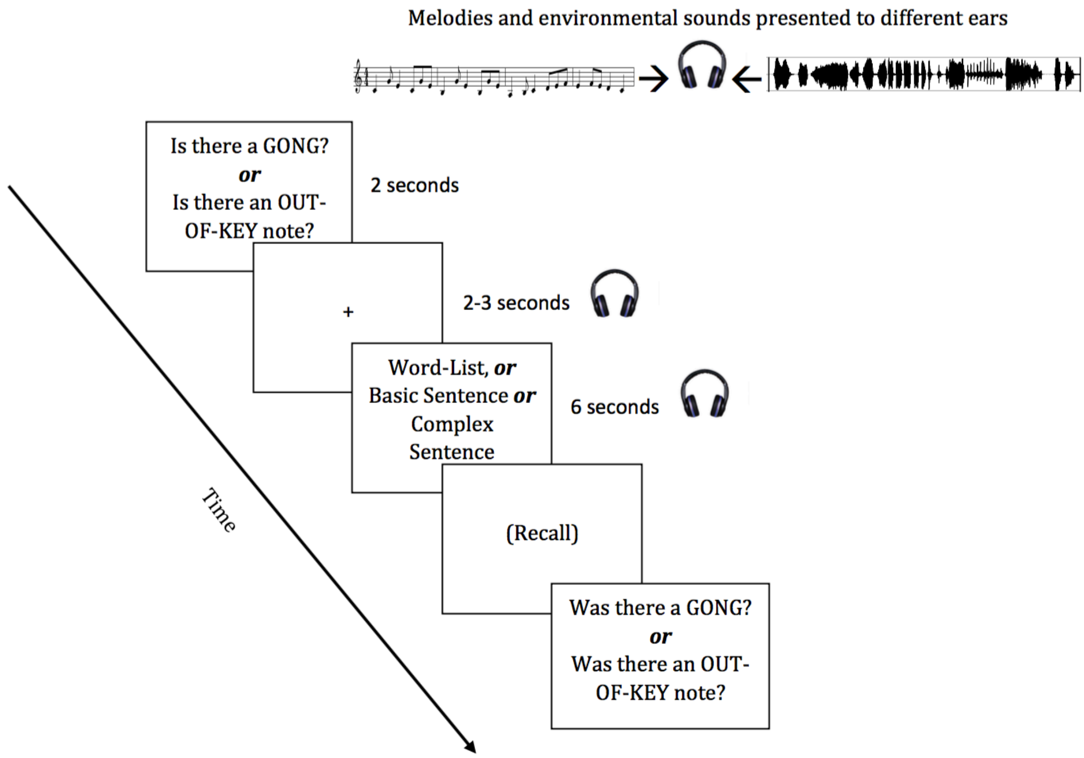

Figure 5. Timeline of stimulus presentation for Experiment 3. Headphones represent auditory stimulus presentation.

Analyses. First, we determined that participants were detecting the targets to the same level in the melodic and environmental stimuli by comparing target detection accuracy for out-of-key notes and gong sounds with a paired-samples *t*-test. This analysis ensured there was no difference in task difficulty.

Second, we analysed the outcomes for the sentences (simple and complex) separately to word-lists for the same reasons as outlined in Experiment 1. Language recall was scored the same as in Experiment 1, and only trials where there was no target were analysed. We ensured that there was no difference in word-list recall depending on attention by conducting a paired-samples *t*-test on the error scores for word-list recall depending on where attention was directed (to melody, to environmental sounds). We predicted there would be no difference in word-list recall between conditions because the word-lists did not contain syntax, and therefore would not show syntactic interference.

Third, to investigate the effects of attention on sentence processing, we conducted a 2 (sentence type: basic, complex) by 2 (attention: melodies, environmental sounds) repeated measures ANOVA. We predicted a significant main effect of attention because attention to the melody should engage syntactic processing resources to a greater extent than attention to environmental sounds, therefore leading to a greater interference with sentence processing. We also predicted a main effect of sentence type, and an interaction effect between sentence type and attention, as previous work suggests that (a) complex sentences are more difficult to process than basic sentences (Staub, 2010) and (b) syntactic interference effects are more likely under higher syntactic demands (Patel, 2008).

Fourth, in an exploratory analysis, we investigated whether there was an effect of musical training (untrained = zero years of private music lessons, $n = 14$; trained ≥ 5 years private music lessons, $n = 21$) on the ability to recall language using a 2 (sentences: basic, complex) by 2 (attention: melodies, environmental sounds) by 2 (musical training: untrained, trained) mixed-design ANOVA. Based on results of Experiment 1 and the literature, we predicted a between-subjects effect of musical training. There were no other empirical or theoretical bases from which to derive additional hypotheses concerning musical training.

A final analysis was conducted to investigate whether there was an effect of language type (word-lists, basic sentences, complex sentences) on the target detection task. For this analysis, we calculated detection accuracy for out-of-key notes in melodies, and gongs in environmental sounds, across each language condition. Detection accuracy was calculated from the four trials with out-of-key notes for each language condition, and the four trials with gongs for each language condition, for each participant. We predicted that participants would be worse at detecting out-of-key notes when reading sentences compared to word-lists, as both tasks draw on syntactic processing resources. We predicted that there should be no difference in gong detection depending on language condition.

Results

Target detection. Accuracy ratings for target detection in the melody and environmental sound streams showed that participants were detecting targets at a similar rate. There was no difference in out-of-key note detection ($M_{\text{accuracy}} = .83$, $SD = .15$) compared to gong detection ($M_{\text{accuracy}} = .82$, $SD = .13$), $t(48) = .06$, $p = .95$, suggesting the two tasks were at a similar difficulty level.

Word-lists. A paired-samples *t*-test showed no difference in word-list error scores when participants were attending to the melody ($M = 1.23$, $SD = .54$) compared to when they were attending to environmental sounds ($M = 1.28$, $SD = .58$), $t(48) = .97$, $p = .34$. As predicted, this finding showed no effect of attention (directed to melodies or to environmental sounds) on word-list recall.

Sentences. A 2 x 2 repeated measures ANOVA with the factors of attention (to melodies, to environmental sounds), and language (basic sentences, complex sentences) showed a significant main effect of attention, $F(1, 48) = 5.44$, $p = .02$, $\eta^2 = .10$, no main effect of language $F(1, 48) = 1.48$, $p = .23$, and no interaction effect $F(1, 48) = .50$, $p = .49$. These findings suggest that the direction of attention to the melody or to environmental sounds had an effect on sentence recall regardless of whether the sentence was complex or basic, though the main effect of attention does appear to be driven by the complex condition. See Figure 6.

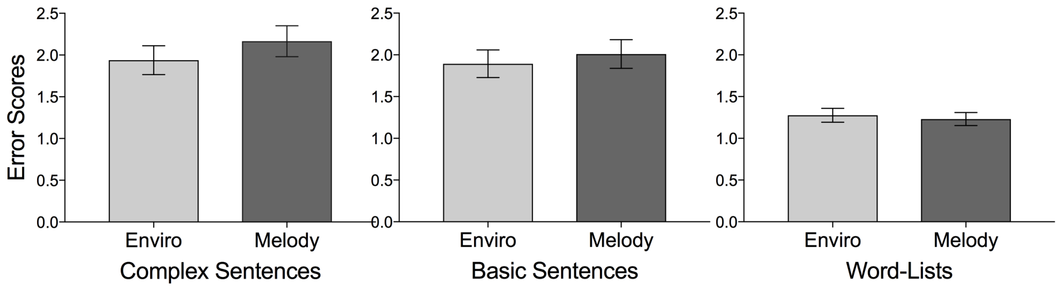

Figure 6. Errors in recall for complex sentences, basic sentences, and word-lists, based on whether participants were attending to either the environmental (enviro) sounds (lighter bars), or the melodies (darker bars). Error bars indicate one standard error either side of the mean.

Musical training. The above ANOVA was conducted on a subset of participants to include the between-subjects factor of musical training (untrained, trained) in an exploratory analysis. The main effect of attention was still significant, $F(1, 33) = 4.26$, $p = .047$, $\eta^2 = .11$. No other main effects or two-way interaction effects were significant, and there was no between-subjects effect of musical training[3], $F(1, 33) = .10$, $p = .76$. However, there was a significant three-way interaction between

[3] There was also no between-subjects effect of which ear melodies were presented to, $F(1, 47) = .10$, $p = .755$.

attention, language, and musical training, $F(1, 33) = 6.56$, $p = .015$, $\eta^2 = .17$. Specifically, participants without musical training made the most errors recalling basic sentences when they were paired with melodies ($M = 2.2$, $SD = 1.1$), whereas musically trained participants made the most errors when recalling complex sentences paired with melodies ($M = 2.05$, $SD = 1.3$). Furthermore, participants without musical training made the least amount of errors recalling basic sentences paired with environmental sounds ($M = 1.71$, $SD = .95$), whereas participants with musical training made the least amount of errors recalling complex sentences paired with environmental sounds ($M = 1.78$, $SD = 1.17$). This result is counterintuitive, as we expected recall for basic sentences to be better than recall for complex sentences in general. Independent-samples *t*-tests showed no differences between the musical training groups on any of the measures, suggesting this interaction is only revealed when taking into account direction of attention, language type, and musical training.

Effects of language on target detection. To investigate whether the type of language participants were reading (word-list, basic sentence, complex sentence) affected their ability to detect out-of-key notes in the melody stream and gongs in the environmental sound stream, we analysed target detection based on language condition. A 3 (language type: word-lists, basic sentences, complex sentences) x 2 (target type: out-of-key note, gong) repeated measures ANOVA was conducted. This analysis revealed a main effect of language type, $F(2, 96) = 7.36$, $p = .001$, $\eta^2 = .13$, no main effect of target type, $F(1, 48) = .003$, $p = .95$, and a marginal but non-significant interaction, $F(2, 96) = 2.48$, $p = .089$. Paired-samples *t*-tests were conducted to examine the main effect of language type. These *t*-tests compared out-of-key note detection and gong detection across the three language conditions (Holm-Bonferroni adjusted *p* values reported for six comparisons).

As can be seen in Figure 7, language type affected out-of-key note detection, but not gong detection. Participants detected out-of-key notes significantly better when concurrently reading word-lists (M = .90, SD = .15) compared to reading both basic sentences (M = .81, SD = .19), $t(48) = 2.9$, p' = .025, d = .41, and complex sentences (M = .76, SD = .20), $t(48) = 4.38$, $p' < .001$, d = .63. Out-of-key note detection did not differ between basic and complex sentences, $t(48) = 1.70$, p' = .38. There were no differences in gong detection between any of the language conditions. Gongs were detected equally when participants were reading word-lists (M = .84, SD = .23) compared to both basic sentences (M = .83, SD = .20), $t(48) = .26$, p' = 1.0, and complex sentences (M = .81, SD = .21), $t(48) = .93$, p' = 1.0. There was also no difference in gong detection when reading basic or complex sentences, $t(48) = .78$, p' = 1.0. This pattern of results suggests that out-of-key note detection is affected by syntax in sentences, but gong detection is not, providing further evidence for shared syntactic processing resources.

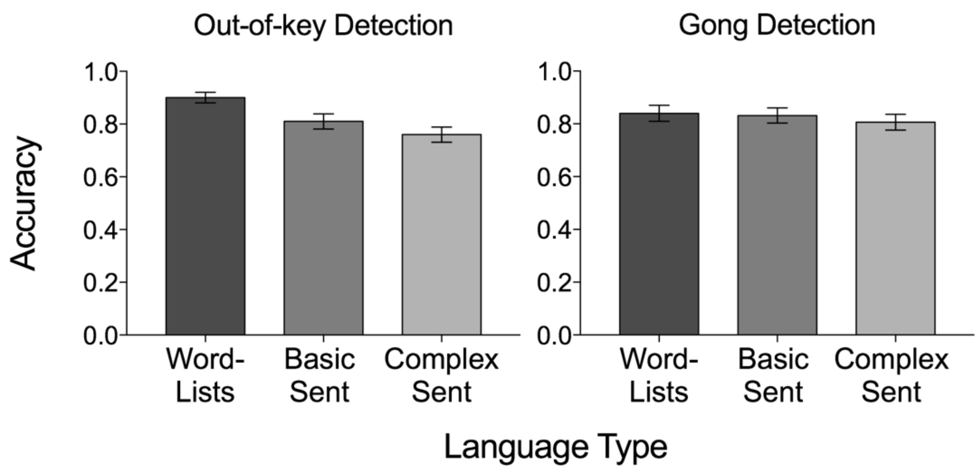

Figure 7. Target detection for out-of-key notes and gong sounds in stimuli depending on concurrent language task.

Discussion

Experiment 3 suggested that attention plays an important role in syntactic processing, corroborating previous research (Loui et al., 2005; Maidhof & Koelsch, 2011). With identical auditory input, sentence recall was significantly affected by where participants were directing their attention—either to a melodic stream containing syntax, or to a stream of environmental sound chunks containing no syntax. When attention was directed to the melodic stream, there was evidence of a syntactic interference effect in sentence recall that was significantly greater than when participants were attending to the environmental sound stream. Although the observed effect size was relatively small, the presence of an effect of attention with identical auditory input is informative. Future research could aim to increase statistical power in order to verify this effect. There was no difference in word-list recall depending on attention, suggesting that this finding is syntax-specific. Detection of out-of-key notes was also affected by sentence type, with participants detecting out-of-key notes significantly better when reading word-lists compared to when reading both basic and complex sentences. This pattern was not found for gong detection, suggesting a syntax-specific effect. Experiment 3 therefore revealed that attention influences the engagement of syntactic processing resources, thereby altering syntactic interference effects.

General Discussion

The results of this series of experiments suggest that syntactic interference effects can be elicited without syntactic violations, and that these interference effects are modulated by attention. These findings provide support for theories of shared syntactic processing between music and language, and suggests that the concurrent processing of music and language reveals interference at a higher level than shared error detection mechanisms.

In Experiment 1, we found that memory for complex sentences (but not word-lists) was significantly decreased when participants were concurrently listening to normal (one timbre) melodies compared to environmental sounds. Interestingly, melodies with alternating timbres resulted in *better* sentence recall than normal melodies, though participants still performed better in the environmental sound condition. This finding was at first surprising, as the alternating timbre condition was designed to be more distracting than the normal melody condition. We hypothesised that alternating timbres disrupted auditory streaming, leading to a weaker syntactic representation, and therefore less interference with language syntax processing.

Experiment 2 confirmed this hypothesis, showing that melodic same-different judgements were more accurate when melodies were played with one instrument compared to three instruments. This result aligns with the changing-state hypothesis (e.g., stimuli that change state take more resources to process; Hughes et al., 2013), Gestalt grouping principles (e.g., alternating timbres were grouped as coming from separate sources; Deutsch, 2013; Zeamer & Fox Tree, 2013), and auditory streaming (e.g., the changing timbres disrupted auditory streaming; Bregman, 1990). Combining the results from Experiment 1 and 2 suggests that interrupting auditory stream formation with alternating timbres leads to impaired syntactic structure building, and therefore a weaker syntactic representation (Bregman, 1990; Koelsch, 2011). Taken together, Experiments 1 and 2 also suggest that syntactic interference effects are distinct from basic "distraction" effects. If they were purely distraction effects, we should see an additive effect of alternating timbres and syntactic interference. The finding that language recall was better in the timbre melody condition than the normal melody condition in Experiment 1 suggests that the timbre melody condition interfered less with language processing.

In Experiment 3, we measured the effect of attention to different auditory streams on language processing. Participants were presented with melodies in one ear and environmental sound streams in the other simultaneously. While attending to only one stream, participants had to process a word-list (no syntax), basic sentence (simple syntax), or complex sentence (complex syntax). Experiment 3 revealed a larger syntactic interference effect (more errors in sentence recall) when participants were attending to melodies compared to environmental sounds, suggesting that attention modulated the syntactic interference effect and engagement of syntactic processing resources. Furthermore, to direct attention, participants were asked to detect either an out-of-key note in the melodies, or a gong sound in the environmental sounds. Participants were significantly worse at detecting out-of-key notes when they were concurrently reading a sentence compared to when they were reading a word-list. This effect did not occur for gong detection, suggesting that sentences interfered with detection of syntactic violations in music, but not detection of gong sounds in an environmental, non-syntactic sequence.

The results from Experiments 1, 2, and 3 help to elucidate the connections between auditory streaming and syntactic processing. According to models of music perception (Koelsch, 2011), feature extraction and grouping (auditory streaming) is a necessary stage before syntactic structure building. Therefore, a disruption to auditory streaming through alternating timbres should result in reduced syntactic processing. Relatedly, Bregman (1990) has suggested a distinction between what he calls *primitive* segregation processes that are involved in initial streaming, and *voluntary* (schema-based) attention processes. He suggested that primitive scene analysis and streaming are based on low-level features (e.g., timbre cues), and that higher-order and schema-based top-down processing is engaged with attention. Bregman (1990) further suggested that although primitive processes occur outside the scope of attention, attention “selects” incoming streams to process at a deeper level. This process can explain the findings of Experiment 3: that attention to the melodic syntax stream led to greater interference effects in language processing. Syntax therefore appears to be a level of processing that is reliant on low-level auditory scene analysis and segregation, and is influenced by higher-level processes involved with attention.

An interesting finding from Experiment 1 was that the scrambled condition led to the poorest performance in both sentence and word-list recall. There are two possible reasons for these results. First, it is possible that the scrambled condition was more distracting in general (owing to its unpredictable temporal and melodic nature). The finding that scrambled melodies resulted in significantly poorer recall than the other auditory conditions for word-lists supports this interpretation. In addition, difference scores for sentence minus word-list recall showed no difference between the scrambled and normal conditions, suggesting a distraction as opposed to syntactic effect. Second, it is also possible that the non-conforming and unexpected nature of the scrambled condition resulted in increased syntactic processing. Complexity measures (taking into account pitch and duration information) showed that the scrambled condition was more “complex” than the normal condition (Eerola, 2016). Future research should try to tease apart the distinction between distraction, complexity, and syntactic complexity, to ensure studies that increase syntactic complexity are in fact measuring syntactic interference, and not just greater engagement of general processing resources.

The current research is the first to show that syntactic interference effects can be observed in stimuli without syntactic violations, a finding that is relevant to background music research. The effects of background music on language processing show mixed results in the literature. For example, some studies that presented participants with popular background music (with lyrics) found detrimental effects on reading comprehension (Anderson & Fuller, 2010; Tze & Chou, 2010). The addition of lyrics is likely to add an additional processing load on top of the music in these cases, and could be a reason why Tze and Chou (2010) found no difference in performance with background classical music compared to a baseline silence condition. A study by Cassidy and MacDonald (2007) found detrimental effects on language recall when music was high in arousal (this effect was more apparent for introverts), and similarly Thompson, Schellenberg, and Letnic (2012) found lower scores on reading comprehension tasks when the music was fast and loud compared to other combinations such as soft and slow. A study by Hallam, Price, and Katsarou (2002) found that performance on a memory task (remember a word within a sentence) was worse than the silence condition when music was arousing, aggressive, and unpleasant; however, performance was better than the silence condition when music was calming. Various other studies have found differences depending on musical preference (Johansson, Holmqvist, Mossberg, & Lindgren, 2012), musical ability (Patston & Tippett, 2011), or no differences at all (Jancke & Sandmann, 2010).

In light of the current results, we suggest that attentional control, personality, and certain properties of the background music are important predictors as to whether background music will affect concurrent language processing. Research has shown that attention is a predictor in multi-tasking in general (Konig, Buhner, & Murling, 2005), and the current results suggest that language processing will be less affected if attention can be directed away from music syntax processing and engaged in the language task. Previous literature has suggested that music can have positive effects on cognitive performance if it increases arousal (Schellenberg & Weiss, 2013). Therefore, if music increases arousal to an optimal level while engaging minimal syntactic processing resources, this combination should lead to no effect or a positive effect on concurrent language processing. However, the optimal level of arousal is highly dependent on personality. It has been suggested that introverts have a higher level of cortical arousal in general, and so require less stimulation to attain their optimal level of arousal; whereas extroverts require more stimulation to reach their optimal level of arousal (Eysenck, 1967). This difference has been shown experimentally, as introverts are more affected by background music in a reading comprehension task compared to extroverts (Furnham & Allass, 1999). Therefore, individual factors of attentional control and personality will impact this interaction.

The background music itself and the language-processing task are also important predictors as to whether there will be interference in language processing when listening to background music. If background music does not attract attention (e.g., there are no salient distracting or unexpected events), it will be less likely to create interference with language processing. If the language task is engaging, and/or the person engaging with the language task is able to maintain attention to the task at hand, then there should be less syntactic interference, and therefore higher performance on the language task. Furthermore, instrumental background music should have less of an effect on language processing compared to music with lyrics, as the lyrics are likely to result in additional syntactic processing. Taken together, we suggest that the effects of background music on language processing depend on a number of complex factors.

Conclusion

The results of the three experiments in this study provide important evidence concerning the nature of syntactic processing resources that are shared between music and language, and how these resources are modulated by attentional mechanisms. Establishing interference effects *without* violations of syntax is a crucial experimental finding, as the results cannot be explained merely by shared error-processing mechanisms. Instead, syntactic interference effects in this context can be attributed to music and language drawing on a shared pool of limited-capacity syntactic processing resources, providing support for theories of shared syntactic processing, including the SSIRH and the SEH. Our findings further suggest that syntactic processing is dependent on successful auditory streaming and that attentional mechanisms engage syntactic processing to a greater or lesser degree. These results fit within a larger framework proposing domain-general syntactic processing resources in the brain that are modulated by attention.

Ethics Statement: All studies were approved by the Macquarie University Human Research Ethics Committee (5201500300), and all participants gave written informed consent to participate.

Acknowledgements: This research was supported by the Macquarie University Research Excellence Scholarship (MQRES), awarded to Anna Fiveash. Thank you to Glenn Schellenberg for valuable discussions and comments on an earlier version of this manuscript, Bob Slevc for insightful discussions about the results, and Aniruddh Patel for valuable discussions of the results and for suggesting the melodic complexity analysis.

Chapter 2: Interference Without Violations

References

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex, 21*(7), 1507-1518. doi: 10.1093/cercor/bhq198

Anderson, S. A., & Fuller, G. B. (2010). Effect of music on reading comprehension of high school students. *School Psychology Quarterly, 25*(3), 178-187.

Besson, M., Faita, F., Peretz, I., Bonnel, A., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science, 9*(6), 494-498.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Cassidy, G., & MacDonald, R. A. R. (2007). The effect of background music and background noise on the task performance of introverts and extraverts. *Psychology of Music, 35*(3), 517-537. doi:10.1177/0305735607076444

Croonen, W. L. M. (1994). Effects of length, tonal structure, and contour in the recognition of tone series. *Perception & Psychophysics*, *55*(6), 623-632. doi:10.3758/bf03211677

Deutsch, D. (2013). *The psychology of music* (3rd ed.). San Diego: Academic Press.

Dowling, W. J. (1971). Recognition of inversions of melodies and melodic contours. *Perception & Psychophysics*, *9*(3), 348-349. doi:10.3758/bf03212663

Chapter 2: Interference Without Violations

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341-354. doi:10.1037/0033-295X.85.4.341

Eerola, T. (2016). Expectancy-violation and information-theoretic models of melodic complexity. *Empirical Musicology Review, 11*(1), 1-17. doi:10.18061/emr.v11i1.4836

Eerola, T., & Toiviainen, P. (2004). *MIDI toolbox: MATLAB tools for music research*. Kopijyvä: Department of Music, University of Jyväskylä.

Eysenck, H. (1967). *The biological basis of personality*. Springfield: Thomas.

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language, 54*(4), 541-553. doi:10.1016/j.jml.2005.12.006

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition, 37*(1), 1-9. doi:10.3758/MC.37.1.1

Fiebach, C., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research, 30*(3), 321-338.

Chapter 2: Interference Without Violations

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music, 42*(2), 190-209. doi:10.1177/0305735612463949

Furnham, A., & Allass, K. (1999). The influence of musical distraction of varying complexity on the cognitive performance of extroverts and introverts. *European Journal of Personality, 13*(1), 27-38. doi:10.1002/(SICI)1099-0984(199901/02)13:1<27::AID-PER318>3.0.CO;2-R

Hallam, S., Price, J., & Katsarou, G. (2002). The effects of background music on primary school pupils' task performance. *Educational Studies, 28*(2), 111-122. doi:10.1080/03055690220124551

Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: Syntactic and semantic structures. *Frontiers in Psychology, 2*, 112. doi:10.3389/fpsyg.2011.00112

Hocking, J., Dzafic, I., Kazovsky, M., & Copland, D. A. (2013). NESSTI: Norms for environmental sound stimuli. *PLoS One, 8*(9), e73382. doi:10.1371/journal.pone.0073382

Hughes, R. W., Hurlstone, M. J., Marsh, J. E., Vachon, F., & Jones, D. M. (2013). Cognitive control of auditory distraction: Impact of task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 539-553.

Jancke, L., & Sandmann, P. (2010). Music listening while you learn: No influence of background music on verbal learning. *Behavioral and Brain Functions, 6*(3). doi:10.1186/1744-9081-6-3

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage, 47*(2), 735-744. doi:10.1016/j.neuroimage.2009.04.090

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience, 20*(11), 1940-1951.

Johansson, R., Holmqvist, K., Mossberg, F., & Lindgren, M. (2012). Eye movements and reading comprehension while listening to preferred and non-preferred study music. *Psychology of Music, 40*(3), 339-356. doi:10.1177/0305735610387777

Jones, D. M., Hughes, R. W., & Macken, W. J. (2010). Auditory distraction and serial memory: The avoidable and the ineluctable. *Noise Health, 12*(49), 201-209. doi:10.4103/1463-1741.70497

Koelsch, S. (2011). Towards a neural basis of music perception - A review and updated model. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00110

Koelsch, S. (2013). *Brain and music*. Oxford, UK: John Wiley & Sons.

Koelsch, S., Gunter, T., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and music: An ERP study. *Journal of Cognitive Neuroscience, 17*(10), 1565-1577.

Konig, C. J., Buhner, M., & Murling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance, 18*(3), 243-266. doi:10.1207/s15327043hup1803_3

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*(4), 334-368.

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS One, 10*(11), e0141069. doi:10.1371/journal.pone.0141069

Kunert, R., Willems, R. M., & Hagoort, P. (2016). Language influences music harmony perception: Effects of shared syntactic integration resources beyond attention. *Royal Society Open Science, 3*(2). doi:10.1098/rsos.150685

Lamont, A. (2016). Musical development from the early years onward. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed.). Oxford: Oxford University Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Levitin, D. J., & Menon, V. (2003). Musical structure procesed in "language" areas of the brain: A possible role for Brodmann area 47 in temporal coherence. *NeuroImage, 20*, 2142-2152. doi:10.1016/S1053-8119(03)00482-8

Loui, P., Grent-'t-Jong, T., Torpey, D., & Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Cognitive Brain Research, 25*(3), 678-687. doi:10.1016/j.cogbrainres.2005.08.019

Maidhof, C., & Koelsch, S. (2011). Effects of selective attention on syntax processing in music and language. *Journal of Cognitive Neuroscience, 23*(9), 2252-2267.

Chapter 2: Interference Without Violations

Mann, V., Liberman, I., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. *Memory & Cognition, 8*(4), 329-335. doi:10.3758/BF03198272

Mathworks (R2014b). Matlab. Massachusetts, United States: Mathworks

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

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733.

Patston, L. M., & Tippett, L. J. (2011). The effect of background music on cognitive performance in musicians and nonmusicians. *Music Perception, 29*(2), 173-183. doi:10.1525/mp.2011.29.2.173

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences, 999*(1), 58-75. doi:10.1196/annals.1284.006

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*(7), 688-691.

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin & Review, 20*(2), 310-317. doi:10.3758/s13423-012-0344-5

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review, 21*(2), 301-308. doi:10.3758/s13423-014-0595-4

Saffran, J. R., Senghas, A., & Trueswell, J. C. (2001). The acquisition of language by children. *Proceedings of the National Academy of Sciences, 98*(23), 12874-12875. doi:10.1073/pnas.231498898

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *Psychology of Music* (3rd ed.). San Diego: Academic Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science*, *22*(11), 1359-1366. doi:doi:10.1177/0956797611417632

Singer, N., Jacoby, N., Lin, T., Raz, G., Shpigelman, L., Gilam, G., . . . Hendler, T. (2016). Common modulation of limbic network activation underlies musical emotions as they unfold. *NeuroImage, 141*, 517-529. doi:10.1016/j.neuroimage.2016.07.002

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review, 16*(2), 374-381. doi:10.3758/16.2.374

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition, 116*(1), 71-86. doi:10.1016/j.cognition.2010.04.002

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18*(5), 1169-1178. doi:10.1093/cercor/bhm149

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music, 40*(6), 700-708. doi:10.1177/0305735611400173

Tillmann, B., & Bigand, E. (2015). A commentary on “A commentary on: ‘Neural overlap in processing music and speech’”. *Frontiers in Human Neuroscience, 9*. doi:10.3389/fnhum.2015.00491

Tze, P., & Chou, M. (2010). Attention drainage effect: How background music effects concentration in Taiwanese college students. *Journal of Scholarship of Teaching and Learning, 10*(1), 36-46.

Vuvan, D. T., Prince, J. B., & Schmuckler, M. (2011). Probing the minor tonal hierarchy. *Music Perception, 28*(5), 461-472.

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings. *Psychological Science*, *23*(10), 1074-1078. doi:doi:10.1177/0956797612442552

Zeamer, C., & Fox Tree, J. (2013). The process of auditory distraction: Disrupted attention and impaired recall in a simulated environment. *Journal of Experimental Psychology: Learning, Memory & Cognition, 39*(5), 1463-1472.

Chapter 3

Syntactic Processing in Music and Language: Effects of Interrupting Auditory Streams With Alternating Timbres

One of the findings from Chapter 2 was that syntactic processes were less engaged for melodies played with multiple musical instruments than for melodies played with a single instrument. This conclusion was inferred from behavioural observations that (a) participants performed better on a same-different melodic judgement task when melodies contained one timbre compared to three timbres; and (b) linguistic processing was less affected by accompanying music if melodies alternated between multiple instruments than if they consisted of a single instrument. The first observation provides direct evidence that alternations in timbre disrupt music syntax processing; the second observation provides indirect evidence for the conclusion, as it implies that enhanced linguistic processing occurred in the three-timbre condition because syntactic processes were not otherwise occupied by the music.

Chapter 3 was designed to corroborate this behavioural evidence in an electrophysiological study, and further investigate Question 2—does disrupting auditory streaming processes influence syntactic processing, and is the connection between auditory streaming and syntactic processing similar for music and language? Event-related potentials (ERPs) were obtained to measure neural responses to syntactic violations in music and language when timbre was manipulated. We presented participants with melodies and sentences that contained either one timbre (one instrument, one voice) or three timbres (three instruments, three voices). This design allowed us to measure ERPs reflecting the processing of syntactic violations, and to observe whether alternating timbres reduced this response.

Throughout Chapter 2, we argued that alternating timbres disrupted syntactic processing because of the critical role that timbre plays in auditory stream segregation. Timbre is used as a perceptual grouping cue to classify incoming sounds from various sources into distinct auditory streams, given that sounds with different timbres are likely to originate from different sources (Bregman, 1990). Therefore, by disrupting auditory streaming with alternating timbres, it should be difficult for listeners to form a coherent representation of melodic syntax. This suggestion is compatible with auditory streaming research (Bregman, 1990), Gestalt grouping principles (Deutsch, 2013a), and models of music and language processing suggesting that feature extraction and auditory streaming is a prerequisite for syntactic processing (Friederici, 2002; Koelsch, 2011, 2013). Chapter 3 tested whether ERP responses to syntactic violations in melodies and spoken sentences were significantly reduced when the stimuli contained three, alternating timbres compared to one timbre.

This manuscript was co-authored by myself, Bill Thompson, Nicholas Badcock, and Genevieve McArthur. I contributed approximately 80% of the total work, including experimental design, stimuli creation, experiment programming, data collection, statistical analysis, and the preparation of the first draft of the manuscript. Bill Thompson and Genevieve McArthur contributed to experimental design, gave critical feedback throughout the process and helped with interpretation of data. Nicholas Badcock helped with experiment programming, data analysis, and data visualisation. All authors provided helpful comments on the manuscript.

Chapter 3: ERP Study

This chapter has been submitted as:

Fiveash, A., Thompson, W. F., Badcock, N. A., & McArthur, G. (under review). Syntactic processing in music and language: Effects of interrupting auditory streams with alternating timbres.

Syntactic Processing in Music and Language: Effects of Interrupting Auditory Streams with Alternating Timbres

Anna Fiveash[1,3], William F. Thompson[1,3], Nicholas A. Badcock[2,3], & Genevieve McArthur[2,3]

[1]Department of Psychology, Macquarie University

[2]Department of Cognitive Science, Macquarie University

[3]ARC Centre of Excellence in Cognition and its Disorders, Macquarie University

Chapter 3: ERP Study

Abstract

Both music and language rely on the processing of spectral (pitch, timbre) and temporal (rhythm) information to create structure and meaning from incoming auditory streams. Previous behavioural results have shown that interrupting a melodic stream with unexpected changes in timbre results in reduced syntactic processing. Such findings suggest that syntactic processing is conditional on successful streaming of incoming sequential information. The current study used event-related potentials (ERPs) to investigate whether (a) the effect of alternating timbres on syntactic processing is reflected in a reduced brain response to syntactic violations, and (b) the phenomenon is similar for music and language. Participants listened to melodies and sentences with either one timbre (piano or one voice) or three timbres (piano, guitar, and vibraphone, or three different voices). Half the stimuli contained syntactic violations: an out-of-key note in the melodies, and a phrase-structure violation in the sentences. We found smaller ERPs to syntactic violations in music in the three-timbre compared to the one-timbre condition, reflected in a reduced early right anterior negativity (ERAN). A similar but non-significant pattern was observed for language stimuli in both the early left anterior negativity (ELAN) and the left anterior negativity (LAN) ERPs. The results suggest that timbre disruptions to auditory streaming reduce syntactic processing for music.

Keywords: music, language, syntax, timbre, attention, ERAN.

Chapter 3: ERP Study

Music and language share similarities in both lower-level perceptual features and higher-level structural features. Lower-level features include changes in pitch, timbre, timing, and intensity, which are fundamental characteristics of music and language (e.g., notes, phonemes). Higher-level features emerge when smaller elements are combined through processes of auditory streaming, which form larger sequences such as musical phrases in music, or linguistic phrases in language. Musical and linguistic phrases are characterized by *syntactic structure*—a system of regularities in how elements are combined (Patel, 2008). Syntax can include hierarchical (nested) structure and strong dependencies between elements (Koelsch, 2013; Patel, 2003, 2008). Implicit knowledge of syntax results in expectations about upcoming events (Huron, 2008).

Although the elements of music and language are different (e.g., notes and chords versus words), there are parallels in how the two domains are processed in the brain. Models of music perception (Koelsch, 2011) and auditory sentence processing (Friederici, 2002) propose similar processing stages, and it has been suggested that music and language may draw upon shared resources for processing syntactic structure (Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009; Fiveash & Pammer, 2014; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Patel, 2003; Sammler et al., 2013; Steinbeis & Koelsch, 2008). The neurocognitive model of music perception (Koelsch, 2011) suggests that musical feature extraction (including pitch, timbre, and intensity information) occurs within the first 100 milliseconds (ms) after stimulus onset. The neurocognitive model of auditory sentence processing (Friederici, 2002) also contains an early feature extraction stage termed *primary acoustic analysis* that occurs within the first 100ms after stimulus onset—before identification of word category and syntactic structure building. These early feature extraction stages directly feed into processes of auditory scene analysis (Bregman, 1990).

Auditory scene analysis is the process by which incoming acoustic information is streamed into meaningful units (Bregman, 1990). The incoming information is grouped into an auditory stream based on Gestalt principles that identify the source of the sound. Sounds that are *similar* (e.g., in timbre) or *proximal* (e.g., in pitch) are mostly grouped within the same auditory stream, as they are quite likely to come from the same source (Bregman, 1990; Deutsch, 2013b; Iverson, 1995; McAdams, 2013). Auditory streaming is suggested to be a prerequisite for later, higher level processes, such as developing a syntactic representation of incoming information (Koelsch, 2013). If auditory streaming were disrupted, it would follow that syntactic processing would also be impaired. Given the important role of timbre in auditory scene analysis, a sequence of multiple, unpredictable timbres are unlikely to be grouped as part of the same auditory stream (Bregman, 1990). This disruption to auditory streaming would in turn have an impact on syntactic processing. Thus, timbre can be used as a tool to investigate processes of auditory streaming and syntactic processing in music and language.

Links between timbre and syntax have been observed in previous research. For example, McAdams (1999) presented participants the same piece of music with either a sampled orchestra (multiple timbres) or a sampled piano (one timbre). McAdams (1999) stopped the music at 23 distinct points, and asked participants at each point to rate how “complete” the music sounded. Lower ratings of completion imply higher tonal tension, whereas higher ratings of completion suggest lower tonal tension (more relaxation). Ratings of completion were significantly higher for the orchestral version of the piece than for the piano version, suggesting that timbre influenced sensitivity to syntactic structure. As musical tension and relaxation patterns are integral to syntactic structure (Huron, 2008), it appears that participants were less sensitive to tension and resolution patterns in the syntax when they were listening to multiple instruments. Cusack and Roberts (2000) further showed that changing timbres in a rhythm discrimination task resulted in poorer performance on a task requiring stream integration, as changing timbres disrupted this process.

Recent behavioural work has suggested that syntactic processing is weakly engaged by melodic sequences that contain alternating timbres (Fiveash, McArthur, & Thompson, under review). In Experiment 1, participants listened to melodies played with one timbre or with three timbres (among other conditions) while recalling complex sentences or word-lists. Contrary to the hypothesis, participants were better able to recall complex sentences when they were accompanied by melodies with three timbres. This finding was surprising, considering changing-state stimuli tend to be more distracting (Jones, Hughes, & Macken, 2010). The authors suggested that the frequent changes in timbre may have interrupted auditory streaming of melodic information, leading to less coherent sequences and weaker syntactic representations. The reduced syntactic processing of the melody was suggested to have left more syntactic resources available for language syntax processing, resulting in reduced interference for sentence recall.

To examine this hypothesis, Fiveash et al. (under review) conducted a second experiment in which they asked participants to compare two sequential melodies. Participants were worse at discriminating between melodies that contained changes in instrument timbre than melodies comprised of a single timbre throughout, indicating that changes in timbre made it difficult for listeners to form a stable and coherent mental representation of melodies. Although previous research has shown links between timbre and auditory streaming (Bregman, 1990), and between timbre and syntax (McAdams, 1999), Fiveash et al. (under review) was the first study to show that interrupting an auditory stream with changes in timbre reduces syntactic processing. These findings are consistent with the possibility that timbre affects syntactic processing because of its powerful role in auditory streaming, and makes it a useful tool in the current experiment (Bregman, 1990).

Based on auditory streaming research (Bregman, 1990; Deutsch, 2013b; Iverson, 1995), links between timbre and syntax (Cusack & Roberts, 2000; Fiveash et al., under review; Koelsch, 2013; McAdams, 1999), and parallels between music and language (Fiveash & Pammer, 2014; Jentschke, Koelsch, & Friederici, 2005; Koelsch et al., 2002; Kunert, Willems, Casasanto, Patel, & Hagoort, 2015; Levitin & Menon, 2003; Maess, Koelsch, Gunter, & Friederici, 2001; Masataka, 2009; Patel, 2008), we predicted that participants would be less sensitive to violations of syntactic structure in music and language when the auditory streams were interrupted with alternating timbres. To evaluate this prediction, we used event-related potentials (ERPs) to measure the brain response to syntactic violations in normal and interrupted melody and sentence streams.

ERPs are used to measure the timing of brain responses to various stimuli (Luck, 2014). ERP studies have established that when participants hear an out-of-key note or musical chord (a violation of syntactic structure), an early right anterior negativity (ERAN) ERP component is elicited approximately 170-220ms after stimulus onset (Koelsch, 2013). This component is measured by calculating a difference ERP waveform which represents the difference between the ERP to a syntactic violation in a melody and the ERP to the same point in the same melody with no such violation present. It has been suggested that the ERAN reflects an interruption to initial structure building processes in the brain (Koelsch, 2013). The ERAN is reliably elicited to both out-of-key chords in a sequence, and out-of-key notes within a melody (e.g., Koelsch, Gunter, & Friederici, 2000; Koelsch et al., 2005; Koelsch & Jentschke, 2008; Miranda & Ullman, 2007).

Koelsch and colleagues have established that the ERAN is distinct from the mismatch negativity (MMN)—an early component elicited in oddball paradigms to a physical or abstract feature deviant (Koelsch et al., 2001). One reason for this distinction is that the ERAN is affected by tonal context—the amplitude is directly related to how expected or unexpected a tone or chord is within a current key. In contrast, the MMN is not affected by tonal context. Thus, Koelsch et al. (2001) concluded that the MMN is a response to physical features of a stimulus, and does not reflect sensitivity to the tonal relationships established by a musical key. To further support this conclusion, it has been shown that the MMN is elicited under heavy sedation whereas the ERAN is not (Koelsch, Heinke, Sammler, & Olthoff, 2006), and that the ERAN interacts with early indices of language syntax processing, whereas the MMN does not (Koelsch et al., 2005). The combination of these findings suggests that the ERAN component is related to music syntax processing in the brain.

Syntactic processing in language can also be studied using ERPs. To date, two early ERP components to syntactic violations in language have been identified. The early left anterior negativity (ELAN) has been observed in response to word category violations and early phrase-structure violations, and is evident at approximately 100-300ms post stimulus onset (Friederici, 2002)—a similar time window to the ERAN. A left anterior negativity (LAN) is evident at approximately 300-500ms post stimulus onset, and is found with morpho-syntactic violations, number disagreements, and gender disagreements (Coulson, King, & Kutas, 1998; Friederici, 2002; Gunter, Friederici, & Schriefers, 2000). Previous research has also found the LAN in response to word category violations (e.g., Hagoort, Wassenaar, & Brown, 2003). The LAN is in the same time window as the N400—a component elicited with semantic errors in language. However, research suggests that the LAN and the N400 reflect two separate processes because of a lack of interaction between the two components (Friederici, 2002; Gunter et al., 2000). Thus, it appears that the ELAN and LAN are early indicators of a syntactic violation in language.

The current study investigated syntactic processing in music and language using the ERAN, ELAN, and LAN ERP components, as well as behavioural measures. In particular, we were interested in whether disrupting auditory streams with alternating timbres has analogous effects on neurophysiological and behavioural indices of syntactic processing in the two domains. We first determined whether our stimuli elicited expected brain responses to syntactic violations (the ERAN in music, and the ELAN or LAN in language). Once we identified these components, we then determined whether the response to a syntactic violation was significantly reduced in the three-timbre conditions (disrupted auditory streaming) compared to the one-timbre conditions (intact auditory streaming). A reduced response in the three-timbre conditions would indicate that alternating timbres led to a reduction in syntactic processing. However, no difference would suggest that alternating timbres did not have an impact on syntactic processing at the neurophysiological level. A similar pattern in music and language would indicate that a disruption to auditory streaming reduces syntactic processing in a similar way across both domains. The current investigation is the first to examine the electrophysiological consequences of disrupting auditory streaming with changes in timbre, and how this disruption impacts syntactic processing in music and language.

Method

Ethics

This study was approved by the Macquarie University Human Research Ethics Committee (ref: 5201500300).

Participants

Twenty-three students from Macquarie University participated for course credit. One participant was excluded due to a recording error, leaving 22 participants (M_{age} = 20 years, range: 18-24; 17 females). All participants were native-English speakers, and 21 reported being right handed. Participants had an average of 4.38 years of private music lessons (range: 0-17), and 8.02 years of combined private and informal experience (range: 0-32). Eight participants had five or more years of private musical training. Three participants indicated that they were musicians, 14 indicated they were non-musicians, and 4 considered themselves as "somewhat" a musician (one participant did not respond to this question). Eight indicated that they were currently musically active. All reported listening to music daily, with an average listening time of 124 minutes per day (*SD* = 90 mins).

Design

The experiment was a 2 (stimuli: melodies, sentences) by 2 (timbre: one, three) by 2 (syntax: violation, no violation) within-subjects design. There were eight conditions, with 50 trials in each condition (i.e., a total of 400 trials per participant). The melodies were played with one timbre or three timbres, with a violation (out-of-key note) or no violation. The sentences were spoken by one speaker or three speakers, with a violation (phrase-structure violation) or no violation. Melodies and sentences were presented in separate blocks, and presentation order was counterbalanced across participants. Within blocks, stimulus presentation was randomised to ensure different presentation for each participant, and the same melody or sentence was never presented in a row. Behavioural and ERP data were recorded simultaneously, and participants had a break every 50 trials.

Stimuli

Stimuli were programmed and presented using Matlab (version R2016b; Mathworks) and Psychtoolbox (version 3.0.13; Brainard, 1997).

Melodies. Fifty musical instrument digital interface (MIDI) melodies were composed in MuseScore in the keys of C, G, D, and A major. The melodies were composed by a professional composer (the second author), and simplified for ERP research by the first author. Each melody started and ended on the tonic note of the key to enhance key strength, were 100 beats per minute (bpm), four bars long, in a 4/4 time signature, and averaged 21 notes (range: 18-24 notes). MIDI melodies were then imported into GarageBand. One-timbre stimuli were played on the Steinway grand piano MIDI instrument, and three-timbre stimuli were played with Steinway grand piano, acoustic guitar, and vibraphone MIDI instruments. An external random number generator determined which instrument played each note, and it was ensured that no instrument played more than two notes in a row.

The one- and three-timbre violation conditions contained an out-of-key note. Stimuli were designed so that the "critical note" (out-of-key note) was always in the final two bars, always fell on a strong (one or three) beat, on a full quarter note, and was always preceded by a full quarter note. Therefore, there was always 600ms after note onset to measure the violation response (i.e., the baseline was not corrupted by the onset of a previous note).

Sentences. Sentences were designed for the same four conditions: one-timbre (violation, no violation), and three-timbres (violation, no violation). Thirty sentences from Neville, Nicol, Barss, Forster, and Garrett (1991) were used, and 20 more with a similar structure were created so there were 50 sentences in total. These sentences, each comprising seven or eight words, were all declarative sentences consisting of noun phrases and a possessor (e.g., *Fred's*). These sentences all had a similar structure, such as: *The widow asked for Fred's advice **about** taxes*. Phrase-structure violations were used, as these have been shown to disrupt early syntactic processing, akin to music syntax violations (Koelsch, 2013). To create the phrase structure violation, the "critical word" (always *about* or *of*), was moved to the position after the possessor, such as: *The widow asked for Fred's **about** advice taxes*. For more information about the sentence constructions, please see Neville et al. (1991).

Three Australian, female, native-English speakers were recorded in a sound proof room. The speakers practiced before the recording, and they were instructed to read each sentence with normal prosody, but with gaps after each word that were long enough that the words did not run together. This recording technique allowed for word splicing in Praat (version 5.4.22; Boersma, 2001), and was implemented to minimise overlap between ERPs to successive words. Sentences were then manipulated in Praat to ensure there was always at least 600ms from the onset of one word to the onset of the next word, and at least 100ms of silence before the onset of each word to maximize a stable pre-stimulus baseline.

The same speaker spoke all sentences in the one-timbre condition. To create the three-timbre condition, an external random number generator determined which speaker would speak each word (with the caveat that the same speaker never said two words in a row). Different speakers' voices were then spliced together using Praat to create sentences.

Critical points. Critical points in the stimuli were marked using Praat. For the melodic stimuli, the critical time points were at the onset of the out-of-key note, and the onset of the same note in the matching no violation stimuli. In the sentence stimuli, the critical time points were the onset of the violation word, and the onset of the same word in the matching no violation stimuli. Event markers were sent to the continuous EEG recording at the onset of each trial using a parallel port, and the critical time point was updated offline.

Procedure

Participants were tested in an electrically and acoustically shielded room. Participants signed the information and consent form, filled out a music education and preference questionnaire, and were instructed about the task. To reduce set up time by reducing electrode impedance, the participant's scalp was combed (Mahajan & McArthur, 2010), face and mastoid areas were cleaned, and electrodes were placed on the face and mastoid bones and filled with a conductive gel. The EasyCap with electrodes attached was then secured on the participant's head, and scalp electrodes were filled with conductive gel. Electrode impedances (measured using the Neuroscan Synamps acquisition system and Scan software; Scan 4.3) were adjusted to be below 5 kΩ. This set-up process took approximately 30 minutes.

Participants were instructed that on each trial of the experiment, their task was to decide whether or not there was (a) an out-of-key note in a melody played by a piano or by three different instruments; or (b) a grammatical error in a sentence that was spoken by either one speaker or three different speakers. Participants heard examples of the stimuli. When participants indicated they understood the task and were comfortable, the experiment began. Stimulus presentation was blocked so that participants were presented with all the melodies first, or all the sentences first. A fixation-cross was first presented on the screen for 1 second, and then the stimuli was presented through headphones for its duration. After each trial, participants indicated on the keyboard whether there was a violation (press z) or whether there was no violation (press m). The experiment took approximately 1 hour and 30 minutes, including set-up time.

Behavioural Measures

Behavioural data consisted of participant responses to the question "was there a violation?" for each trial. To analyse these responses, d prime (d') sensitivity scores were calculated to measure how sensitive participants were to out-of-key notes in melodies and grammatical errors in sentences. These values were calculated by subtracting the z scores for each participant's *false alarm* rate (when there was no error and the participant said there was an error) from the *hit* rate z score (when there was an error and the participant detected an error). Extreme values (e.g., 100% or 0% accuracy) were corrected for (Stanislaw & Todorov, 1999). A measure of response bias "c" was also calculated (see Stanislaw & Todorov, 1999) which showed whether participants were more biased towards responding yes or no overall. Positive c scores reflect a bias towards responding no, and negative scores reflect a bias towards responding yes. A score of 0 indicates no bias.

EEG Recording

Electroencephalography (EEG) was recorded using the Neuroscan system (version 4.3) and a Synamps2 amplifier with a sampling rate of 1000 hertz (Hz), and an online bandpass filter (1-100 Hz). Brain activity was measured through 30 electrodes positioned according to the 10-20 system (EasyCap; Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, P8, O1, Oz, O2). The ground electrode was located at AFz, and reference electrodes were placed on the left and right mastoid bones. Horizontal electro-oculographic (HEOG) activity was recorded using electrodes placed on the left and right outer canthi of the eyes. Vertical electro-oculographic (VEOG) activity was recorded using electrodes placed above and below the left eye.

ERP Processing

Data was processed using EEGLAB (version 13; Delorme & Makeig, 2004) and Matlab (version 2016b; Mathworks). EEG recorded from each electrode was filtered with a high-pass filter of 0.1 Hz and a low-pass filter of 30 Hz. The online reference was the left-mastoid (M1), and then the data were re-referenced offline to the right-mastoid (M2), effectively taking the average of the two mastoids. An independent components analysis (ICA) was conducted on all the data from all electrodes in EEGLAB. Eye blink components were removed based on visual inspection of ICA components. Data were then epoched to 700ms after the onset of the critical note or word with a baseline correction of 100ms, resulting in 50 epochs in each condition for each participant. Epochs with extreme values at the sites of interest (frontal left and right electrodes—F7, FT7, F3, FC3, F8, FT8, F4, FC4) outside the range of -150 to 150 microvolts were removed. This resulted in a .8% loss of epochs across all the different conditions across all participants (melodies: one timbre no violation (8 epochs), one timbre violation (12 epochs), three timbres no violation (7 epochs), three timbres violation (12 epochs); sentences: 9, 7, 10, 7 epochs, respectively). Individual participants had between 0-18 epochs ($M = 3.2$, $SD = 4.6$) removed (out of a possible 400). The remaining epochs in each of the eight conditions were then averaged to create ERPs of each participant's response for each condition (for melodies and sentences:

one-timbre no violation, one-timbre violation, three-timbres no violation, three-timbres violation).

ERP Components

The ERAN, ELAN, and LAN ERP components have reliably been observed at anterior sites, and are reflected primarily in the frontal left and right electrodes (Friederici, 2002; Koelsch, 2013; Koelsch et al., 2001; Maidhof & Koelsch, 2011). Therefore, we focused our analyses on the average of the frontal left (F7, FT7, F3, FC3) and frontal right (F8, FT8, F4, FC4) electrodes. Considering our one- and three-timbre conditions had distinct acoustic differences due to the different timbres (instruments and voices), we calculated the difference waves of the violation condition minus the no violation condition for each individual to isolate the response to out-of-key notes in melodies, and phrase-structure violations in sentences. Based on a visual analysis of the ERP components and previous research (Koelsch, 2013), we defined the ERAN time period of interest as 150-250ms for the melodic stimuli. For the sentence stimuli, previous research has suggested that a phrase-structure violation results in an ELAN, reported to be around 100-300ms (Friederici, 2002; Koelsch, 2013). However, research has also shown an anterior negativity (non-lateralised, between 300-500ms) to word category violations, suggesting that our stimuli may also elicit a later negativity (Hagoort et al., 2003). A visual analysis of our data revealed two negative going peaks in the sentence difference waves, which appear to reflect the ELAN at 100-150ms and the LAN (Friederici, 2002; Hagoort et al., 2003; Koelsch, 2013) at 270-360ms. We therefore analysed both time frames in the sentence stimuli.

Within the time periods of interest for melodies and sentences, we extracted the peak negativity, and calculated the 50ms average around this peak (25ms either side) for each individual in each condition, for both hemispheres. We calculated the average around the peak to get a sensitive measure of the brain response within the time periods of interest, tailored to each individual. For melodies, we extracted one peak in the ERAN time window (150-250ms). For sentences, we extracted two peaks—one in the ELAN time window (100-150ms), and one in the LAN time window (270-360ms). The hemisphere with the largest negativity in the selected timeframe was then used in the analysis. We chose to use the hemisphere with the strongest response, as participants in our sample differed in their lateralisation of syntactic violations for both melodies and sentences. Although the ERAN is generally right lateralised (Koelsch, 2013), and the ELAN and LAN are generally left lateralised (though prosody appears to be processed in the right hemisphere; Friederici, 2002), there have also been a number of studies which have shown a bilateral distribution of both the LAN (Hagoort et al., 2003), and the ERAN (Garza Villarreal, Brattico, Leino, Østergaard, & Vuust, 2011; Loui, Grent-'t-Jong, Torpey, & Woldorff, 2005). In addition, (a) the processing of timbre in the brain is not well understood (Reiterer, Erb, Grodd, & Wildgruber, 2008), (b) it is possible that the unusual nature of our three-timbre stimuli may have led to differences in lateralisation between participants (Boucher & Bryden, 1997), and (c) differences in lateralization have also been found for musicians, who tend to show a greater bilateral distribution of the ERAN (Ono et al., 2011). These findings, combined together, suggest that the lateralisation of the ERAN, ELAN, and LAN cannot be presumed in all subjects, and hence we analysed data from the hemisphere with the greatest response for each condition for the component of interest. We will continue to use the naming conventions (ERAN, ELAN, and LAN) for comparison with the literature, but it should be noted that these are not necessarily lateralised.

Results

Behavioural Results

Melodies. D prime sensitivity measures revealed that participants were significantly better at detecting out-of-key notes in the one-timbre condition (*M* = 2.79, *SD* = .45) than the three-timbre condition (*M* = 2.26, *SD* = .77), $t(21) = 5.22$, $p < .001$, *d* = .84, see Figure 1. Both the one-timbre (*M* = .62, *SD* = .37) and three-timbre (*M* = .22, *SD* = .38) conditions showed a bias towards responding no. However, a paired-samples *t*-test found that the one-timbre condition led to a significantly stronger bias towards responding no than the three-timbre condition, $t(21) = 5.29$, $p < .001$, *d* = .62, suggesting that participants were more likely to give false alarms in the three-timbre condition.

Sentences. Sensitivity measures (d') showed no difference in sentence error detection between the one-timbre condition (*M* = 4.06, *SD* = .49) and the three-timbre condition (*M* = 3.98, *SD* = .52), $t(21) = .80$, $p = .43$ (see Figure 1). There was also no difference between the one-timbre (*M* = .05, *SD* = .19) and three-timbre (*M* = .05, *SD* = .24) conditions in the measure of response bias c, $t(21) = .07$, $p = .95$. This finding may be due to ceiling effects, as the grammatical errors were very obvious, and participants were detecting them with high accuracy.

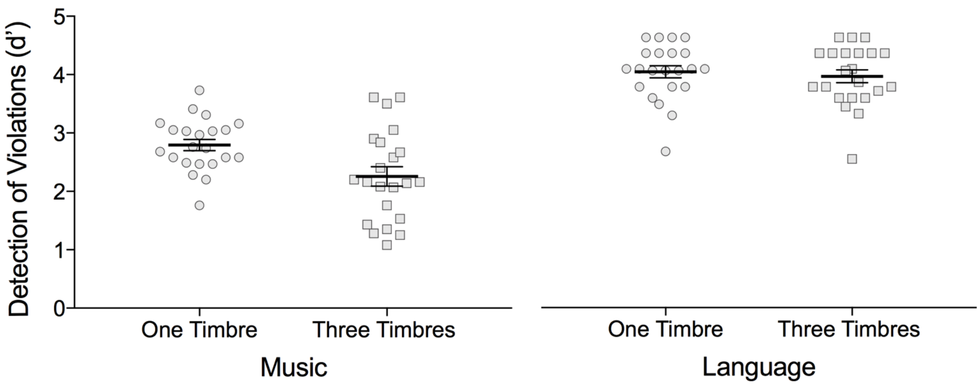

Figure 1. D prime values reflecting sensitivity to out-of-key notes in melodies and grammatical errors in sentences. Individual data points reflect individual participant scores, and the mean is represented by the black line. Error bars indicate one standard error either side of the mean.

ERPs

Reliability of ERAN, ELAN, and LAN. Our first goal was to ensure that our stimuli elicited reliable ERAN, ELAN, and LAN responses in each participant's dominant hemisphere. These data are shown in Figure 2. Means and standard deviations are illustrated in Table 1.

These data confirmed that out-of-key notes within melodies generated an ERAN. One-sample *t*-tests revealed that the ERAN difference wave component was significantly different to zero in the ERAN time window (150-250ms), for both the one-timbre, $t(21) = 6.66$, $p < .001$, $d = 1.42$, and three-timbre, $t(21) = 6.04$, $p < .001$, $d = 1.29$, conditions.

One-sample *t*-tests were also conducted in the ELAN and LAN time windows for the sentence stimuli. Holm-Bonferroni adjusted *p* values are reported for two comparisons in each time window. In the ELAN time window (100-150ms), the one-timbre condition was significantly different to zero, $t(21) = 2.81$, $p' = .022$, $d = .60$, while the three-timbre condition was not, $t(21) = 1.84$, $p' = .08$. In the LAN time window (270-360ms), one-sample *t*-tests confirmed that the difference waves were significantly different to zero for both the one-timbre, $t(21) = 4.65$, $p' < .001$, $d = .99$ and three-timbre, $t(21) = 3.54$, $p' = .002$, $d = .75$, conditions. These findings suggest that (a) the ELAN was evident in the one-timbre condition but not the three-timbre condition, and (b) the LAN was evident in both the one- and three-timbre conditions.

The difference ERP waveform in the LAN time window appeared more reliable than the ELAN response, due to its larger amplitude and existence in both the one- and three-timbre conditions. Thus, the subsequent analyses focused on the LAN rather than the ELAN as a neural index of a phrase-structure violation in language. However, it is interesting to note that the ELAN was evident (though quite weak) for the one-timbre condition, but was not evident for the three-timbre condition. The difference between the one-timbre and three-timbre conditions in the ELAN time window was not significant, $t(21) = 1.24$, $p = .23$.

Effect of disrupting auditory streaming on the ERAN and LAN. To further investigate the effects of alternating timbres on the ERP components related to violations of syntax in melodies and sentences, planned paired-samples *t*-tests were conducted on the 50ms average around the peak of the difference waves (as described above). A paired-samples *t*-test revealed that the response in the ERAN time window to violations of music syntax was significantly more negative in the one-timbre condition compared to the three-timbre condition, $t(21) = 2.74$, $p = .012$, $d = .71$, as predicted.

The sentence data in the LAN time window showed the same pattern of results; however, a paired-samples *t*-test showed no significant difference between the one-timbre condition and the three-timbre condition, $t(21) = 1.04$, $p = .309$. The non-significant difference between the two conditions appears to be because of large variance within our sample, which will be discussed further in the Discussion.

Thus, the response to a music syntax violation in the three-timbre condition was reduced compared to the one-timbre condition, as predicted. Though the same pattern of results was observed in response to a language syntax violation, the difference between the one- and three-timbre conditions was not significant.

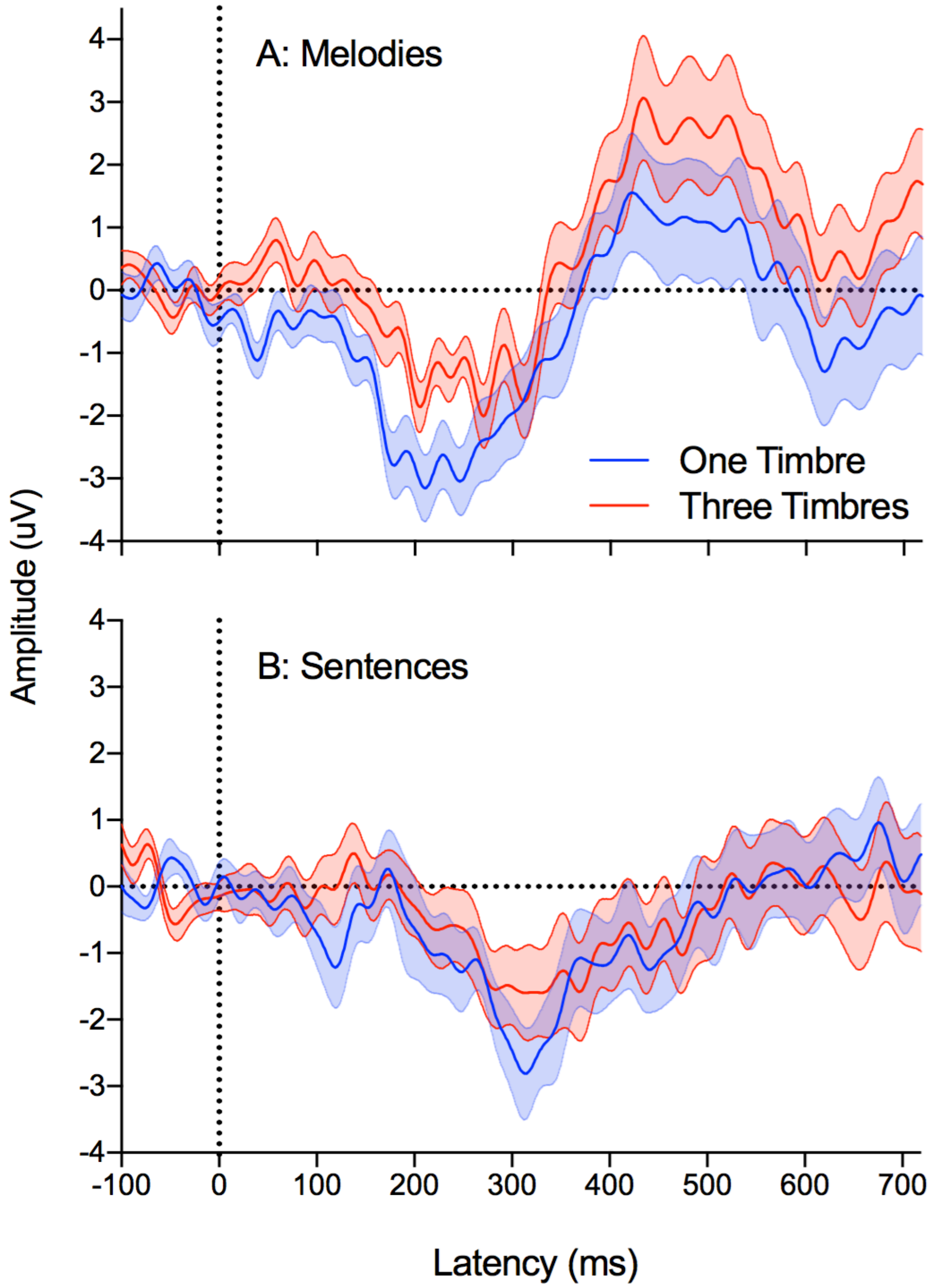

Figure 2. Difference waves for the one-timbre and three-timbre conditions for melodies and sentences. Data is based on the hemisphere with the largest 50ms average around the peak in the time window of interest for (A) Melodies (150-250ms), and (B) Sentences (270-360). Shaded error bars indicate one standard error either side of the mean.

Table 1

ERP Mean Amplitudes and Standard Deviations for the 50ms Average around the Peak in the time window of interest (indicated in brackets)

	One-Timbre		Three-Timbres	
Stimuli	*M*	*SD*	*M*	*SD*
Melodies (150-250ms)	-3.41	2.40	-1.98	1.54
Sentences (100-150ms)	-1.39	2.32	-.62	1.58
Sentences (270-360ms)	-2.91	2.94	-2.16	2.87

Discussion

The current experiment investigated whether behavioural and electrophysiological responses to syntactic violations in melodies and sentences were reduced when syntactic sequences were disrupted with alternating timbres (three-timbre condition) compared to when they were within one auditory stream (one-timbre condition). For melodies, our behavioural data revealed that participants were significantly more sensitive to syntactic violations in the one-timbre condition compared to the three-timbre condition. This finding was also reflected in the ERP results. The ERAN response to out-of-key notes was significantly reduced when melodies were played with three alternating instruments compared to only one instrument. This finding suggests that alternating timbres affect the processing of music syntax in the brain, likely due to an interruption of auditory streaming processes at an early stage. For spoken sentences, we did not observe a significant difference behaviourally or electrophysiologically between the one- and three-timbre conditions, although the LAN ERP response was attenuated in the three-timbre condition compared to the one-timbre condition.

Musical Syntax and Timbre

Previous behavioural research has suggested that alternating timbres in a musical sequence reduces processing of syntactic structure (Fiveash et al., under review; McAdams, 1999). However, the current investigation is the first to investigate this phenomenon with ERPs, which allowed us to investigate the effects of timbre on the neural processing of syntax in real time. To detect syntactic violations, the brain must continuously track incoming information, and register when there is an element that does not adhere to the tonal context. Despite the apparent sophistication of this process, the operation occurs automatically and without overt attention to the stimuli (Loui et al., 2005). In the current experiment, participants exhibited the ERAN in response to out-of-key notes in both the one- and three-timbre conditions, suggesting that the out-of-key note was registered in both conditions. However, this response was significantly reduced when the melodies were played by three timbres compared to one timbre, showing a direct influence of timbre on syntactic processing. The reduced brain response to syntactic violations in the three-timbre condition aligns with our behavioural result that participants were less sensitive to out-of-key notes in the three-timbre condition compared to the one-timbre condition.

The reduced brain response to a syntactic violation when the melody is played with three timbres may be because of the disrupting effect of timbre changes on auditory streaming. Disrupting auditory streaming in turn affects syntactic structure building, resulting in a less coherent melody and a weaker syntactic representation. Perceptual streaming accounts (Bregman, 1990; Cusack & Roberts, 2000; Iverson, 1995) and Gestalt principles (Deutsch, 2013b) have suggested that incoming auditory streams are grouped together by similarity (e.g., timbre) and proximity (e.g., pitch distance). Furthermore, models of music perception and auditory sentence processing include an initial feature extraction and acoustic analysis stage where timbral information is processed (Friederici, 2002; Koelsch, 2011). By disrupting a salient similarity cue (timbre) in early stages of perceptual analysis, and placing a larger burden on auditory streaming processes, it is likely our stimuli made it more difficult for participants to group notes into a coherent stream. However, grouping was not prevented entirely, as participants were able to detect violations in both conditions. It is possible that other grouping principles, such as pitch proximity and regular timing, promoted partial streaming of the melodic information. We therefore suggest that the strength of syntactic representations in the brain is directly related to early auditory streaming processes.

With alternating timbres rendering the melody less coherent, it is possible that predictive processes were also less efficient. Prediction is an important element in both music and language, and can operate on multiple levels (Patel & Morgan, 2016). An out-of-key note in a one-timbre sequence is more unexpected than an out-of-key note in a three-timbre sequence, as the rest of the stream is expected and easily predicted. In a three-timbre context, the timbre of the melodic stream is less predictable, and so the brain may hold weaker predictions about upcoming elements in relation to syntax as well. When these predictions are violated, it may therefore come as less of a surprise. Overall, the current experiment shows that alternating timbres disrupt the processing of syntactic errors in music, at the level of both behaviour and the brain. Interestingly, participants were more likely to give false alarms in the three-timbre melody condition, implying that participants may have been misreading cues to syntactic violations.

It may be valuable for future research to investigate the effects of other methods of disrupting auditory streaming. If the ERAN in the three-timbre condition was reduced because of a disruption to processes of auditory streaming, then any manipulation that disrupts auditory streaming should lead to a similar reduction in the brain response to syntactic violations. Conversely, it is possible that changes in timbre were merely more distracting, and hence drew attentional resources away from auditory streaming processes (Jones et al., 2010). Although possible, this explanation seems unlikely for three reasons. First, an extensive body of research has suggested that the processing of timbre and the formation of auditory streams are not separate processes (Bregman, 1990). Therefore, changes in timbre have a direct impact on the formation and coherence of auditory streams. Second, syntactic processing tends to occur automatically, even when people are not paying attention to the stimulus (though attention does impact this process; Loui et al., 2005). Thus, even if participants were distracted by timbre changes, if the incoming sequences were perceived as coherent streams, then participants should still have had a strong response to the out-of-key note. Third, our previous research revealed that when melodies and sentences were presented concurrently, three-timbre melodies reduced interference for recall of accompanying sentences. If alternating timbres were generally distracting, then we would have expected greater interference by melodies on recall of accompanying sentences (Fiveash et al., under review).

Language Syntax and Timbre

The similarities between music and language in relation to syntax motivated the prediction that three timbres within a sentence (three voices) may also reduce the brain's response to syntactic violations compared to one timbre (one voice). We predicted that the ELAN would be observed in response to phrase-structure violations, as seen in previous literature (Friederici, 2002; Maidhof & Koelsch, 2011). This prediction was only partially supported, as we found a small but statistically reliable ELAN to syntactic violations in sentences in the one-timbre condition but not the three-timbre condition. In contrast, the LAN offered a more reliable response, with a larger peak within the 270-360ms time window in both the one-timbre and three-timbre conditions. Our analysis revealed that the LAN in the three-timbre condition was reduced compared to the one-timbre condition, though this difference was not significant. The lack of a significant difference appears to be due to the large amount of variation between participants.

Finding a statistically reliable effect of timbre changes in music stimuli but not language stimuli was unexpected. There are at least five potential explanations for the difference between the music and language stimuli. First, repeated exposure to conventional Western instruments may have formed an expectation for a high level of consistency in timbre for different events within a musical stream, such that changes in timbre readily disrupt processes of auditory streaming. In contrast, listeners may be more tolerant to changes in vocal timbre within a given speech stream, because speakers routinely use such changes in vocal timbre as part of prosodic communication. More generally, auditory sentence processing is inherently variable, as listeners must process words, prosody, and semantics in addition to syntax, which could have resulted in a noisier signal. Second, in the current experiment, the music stimuli consisted of temporally regular sequences in 4/4 timing, which were therefore highly predictable. Our language stimuli, in contrast, may have sounded rhythmically unnatural, thereby obscuring the effect. Third, it is possible that timbre is more important to syntactic processing for music than for language, as cues to timbre are not as indicative to meaning in language as they are to music. Meaning in language is delivered irrespective of timbre, due to the referential and propositional nature of language (Jackendoff, 2009). On the other hand, meaning in music is a complex phenomenon, related to a number of aspects of the music including pitch and timbre (Koelsch et al., 2004).

Because of this distinction in the way meaning is communicated in music and language, timbral cues may contribute less to the processing of syntax in language than in music.

A fourth consideration is that the grammatical errors were more obvious in the sentence stimuli, as evidenced by ceiling effects in our behavioural data. It is possible that we did not observe a difference between the one- and three-timbre conditions because the task was too easy in comparison with the music task. A fifth possible reason why we did not find any effect in the language condition could be due to our stimuli. Sentence stimuli were created to ensure there was a baseline of silence before the critical word, so that the ERP to the critical word was not affected by the previous word. This manipulation may have resulted in unnatural sounding speech which could have resulted in "noisy" brain activity.

To continue to investigate links between language syntax and timbre, future research could have more trials per condition for a larger signal to noise ratio, create more naturalistic stimuli without disrupting prosody, and introduce more sensitive grammatical errors to try and elicit the effect. For example, obvious syntactic errors may be easily perceived regardless of timbre. It would also be interesting to investigate whether the timbre effect occurs for semantic errors in sentences. If the current study were to be repeated, sentences could be designed so that every word ends on a "stop" consonant (e.g., k, t, p), so that words do not run together. This manipulation would make it easier to splice different voices together without a pause between words.

Conclusion

The current experiment shows, for the first time, that the brain response to syntactic violations in music is reduced when melodies are played by three timbres compared to one timbre. Within a musical perception framework, this finding suggests that alternating timbres disrupt auditory streaming processes in an initial feature extraction stage, which in turn impairs syntactic structure building processes. Although the same pattern was observed for sentence processing, the difference was not significant, likely due to high individual variation in brain responses to auditory sentences. It would be useful if future studies could further explore brain responses to syntactic violations in speech by using carefully controlled stimuli, and increasing the signal to noise ratio. It would also be useful to investigate whether the current findings for the music stimuli can be generalised to different timbres and different methods of disrupting the auditory stream.

Acknowledgements

This research was supported by the Macquarie University Research Excellence Scholarship (MQRES), awarded to the first author.

References

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341-347.

Boucher, R., & Bryden, M. P. (1997). Laterality effects in the processing of melody and timbre. *Neuropsychologia, 35*(11), 1467-1473. doi:10.1016/S0028-3932(97)00066-3

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21-58. doi:10.1080/016909698386582

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics, 62*(5), 1112-1120. doi:10.3758/BF03212092

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21. doi:10.1016/j.jneumeth.2003.10.009

Deutsch, D. (2013a). Grouping mechanisms in music. In D. Deutsch (Ed.), *Psychology of Music*. San Diego: Elsevier.

Deutsch, D. (2013b). *The psychology of music* (3rd ed.). San Diego: Academic Press.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition, 37*(1), 1-9. doi:10.3758/MC.37.1.1

Fiveash, A., McArthur, G., & Thompson, W. F. (under review). Music and language: Syntactic interference without syntactic violations.

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music, 42*(2), 190-209. doi:10.1177/0305735612463949

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84. doi:10.1016/S1364-6613(00)01839-8

Garza Villarreal, E. A., Brattico, E., Leino, S., Østergaard, L., & Vuust, P. (2011). Distinct neural responses to chord violations: A multiple source analysis study. *Brain Research, 1389*, 103-114. doi:10.1016/j.brainres.2011.02.089

Gunter, T., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*(4), 556-568.

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research, 16*(1), 38-50. doi:10.1016/S0926-6410(02)00208-2

Huron, D. (2008). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance, 21*(4), 751-763. doi:10.1037/0096-1523.21.4.751

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception: An Interdisciplinary Journal, 26*(3), 195-204. doi:10.1525/mp.2009.26.3.195

Jentschke, S., Koelsch, S., & Friederici, A. D. (2005). Investigating the relationship of music and language in children: Influences of musical training and language impairment. *Annals of the New York Academy of Sciences, 1060*, 231-242. doi:10.1196/annals.1360.016

Jones, D. M., Hughes, R. W., & Macken, W. J. (2010). Auditory distraction and serial memory: The avoidable and the ineluctable. *Noise Health, 12*(49), 201-209. doi:10.4103/1463-1741.70497

Koelsch, S. (2011). Towards a neural basis of music perception - A review and updated model. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00110

Koelsch, S. (2013). *Brain and music*. Oxford, UK: John Wiley & Sons.

Koelsch, S., Gunter, T., & Friederici, A. D. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12*(3), 520-541.

Koelsch, S., Gunter, T., Schroger, E., Tervaniemi, M., Sammler, D., & Friederici, A. D. (2001). Differentiating ERAN and MMN: An ERP study. *NeuroReport, 12*(7), 1385-1389.

Koelsch, S., Gunter, T., v. Cramon, D., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage, 17*(2), 956-966. doi:10.1006/nimg.2002.1154

Koelsch, S., Gunter, T., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and music: An ERP study. *Journal of Cognitive Neuroscience, 17*(10), 1565-1577.

Koelsch, S., Heinke, W., Sammler, D., & Olthoff, D. (2006). Auditory processing during deep propofol sedation and recovery from unconsciousness. *Clinical Neurophysiology, 117*(8), 1746-1759. doi:10.1016/j.clinph.2006.05.009

Koelsch, S., & Jentschke, S. (2008). Short-term effects of processing musical syntax: An ERP study. *Brain Research, 1212*, 55-62. doi:10.1016/j.brainres.2007.10.078

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. doi:10.1038/nn1197

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS One, 10*(11), e0141069. doi:10.1371/journal.pone.0141069

Levitin, D. J., & Menon, V. (2003). Musical structure procesed in "language" areas of the brain: A possible role for Brodmann area 47 in temporal coherence. *NeuroImage, 20*, 2142-2152. doi:10.1016/S1053-8119(03)00482-8

Loui, P., Grent-'t-Jong, T., Torpey, D., & Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Cognitive Brain Research, 25*(3), 678-687. doi:10.1016/j.cogbrainres.2005.08.019

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). USA: MIT Press.

Maess, B., Koelsch, S., Gunter, T., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4*(5), 540-545.

Mahajan, Y., & McArthur, G. (2010). Does combing the scalp reduce scalp electrode impedances? *Journal of Neuroscience Methods, 188*(2), 287-289. doi:10.1016/j.jneumeth.2010.02.024

Maidhof, C., & Koelsch, S. (2011). Effects of selective attention on syntax processing in music and language. *Journal of Cognitive Neuroscience, 23*(9), 2252-2267.

Masataka, N. (2009). The origins of language and the evolution of music: A comparative perspective. *Physics of Life Reviews, 6*(1), 11-22. doi:10.1016/j.plrev.2008.08.003

Mathworks (R2016b). Matlab. Massachusetts, United States: Mathworks.

McAdams, S. (1999). Perspectives on the contribution of timbre to musical structure. *Computer Music Journal, 23*(3), 85-102. doi:10.1162/014892699559797

McAdams, S. (2013). Musical timbre perception. In D. Deutsch (Ed.), *Psychology of Music*. USA: Elsevier, Inc.

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. doi:10.1016/j.neuroimage.2007.07.034

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*(2), 151-165. doi:10.1162/jocn.1991.3.2.151

Ono, K., Nakamura, A., Yoshiyama, K., Kinkori, T., Bundo, M., Kato, T., & Ito, K. (2011). The effect of musical experience on hemispheric lateralization in musical feature processing. *Neuroscience Letters, 496*(2), 141-145. doi:10.1016/j.neulet.2011.04.002

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674-681.

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

Patel, A. D., & Morgan, E. (2016). Exploring cognitive relations between prediction in language and music. *Cognitive Science*, *41*(S2), 303-320. doi:10.1111/cogs.12411

Reiterer, S., Erb, M., Grodd, W., & Wildgruber, D. (2008). Cerebral processing of timbre and loudness: fMRI evidence for a contribution of Broca's area to basic auditory discrimination. *Brain Imaging and Behavior, 2*(1), 1-10. doi:10.1007/s11682-007-9010-3

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H. J., . . . Schulze-Bonhage, A. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *NeuroImage, 64*, 134-146. doi:10.1016/j.neuroimage.2012.09.035

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18*(5), 1169-1178. doi:10.1093/cercor/bhm149

Chapter 4

Complexity or Syntax? Music, Language, and Syntactic Interference

Chapter 3 confirmed that interrupting an auditory stream with alternating timbres resulted in reduced syntactic processing of a melodic sequence. This finding helps to explain why Experiment 1 in Chapter 2 revealed greater interference for sentence recall by coherent melodies (i.e., with a single timbre) than by melodies involving alternating timbres. Specifically, coherent melodies engaged syntactic processing resources, making these resources less available for concurrent syntactic processing of sentences. In contrast, incoherent melodies (with alternating timbres) only weakly engaged syntactic processing, leaving greater syntactic resources available for sentence processing.

Results reported in Chapter 2 also indicated that recall of sentences and word-lists was poorer when accompanied by scrambled melodies than when accompanied by unscrambled melodies. This outcome was surprising. The process of scrambling melodies was originally designed to eliminate syntactic structure, and hence should not have engaged syntactic processes that could interfere with language syntax processing. However, further analysis in Chapter 2 revealed that the scrambled melodies were significantly more complex than the unscrambled melodies, based on a musical complexity algorithm designed by Eerola (2016). Thus, it is possible that scrambled melodies placed a heavy demand on syntactic processing, as the listeners attempted to make sense of the highly complex melodic sequences. The results from Chapter 2 raise two possibilities: (a) complex melodies place a high burden on music syntax processing which, in turn, interferes with the processing of language syntax, or (b) complex melodies place a high demand on general processing resources, and this general processing demand interferes with the processing of language syntax.

The experiments described in Chapter 4 were conducted to investigate these two possibilities, and to further investigate Question 1—can syntactic interference be observed without violations of syntax? Specifically, the experiments were designed to separate musical complexity and musical syntax processing by investigating whether increases in the complexity of a musical sequence interfere with language comprehension more than increases in the complexity of an environmental sound sequence. Increases in the complexity of music and environmental sounds should place a higher burden on general resources, but only increases in the complexity of music should place a higher burden on syntactic processing. Therefore, Chapter 3 investigated the complexity of alternating timbres and the effect of these timbres on auditory streaming, whereas Chapter 4 investigated complexity of sequences within a single auditory stream.

Participants listened to basic and complex versions of both musical and environmental sound sequences whilst performing either a language comprehension task that relied on accurate syntactic processing without introducing syntactic violations, or a visuospatial search task that did not rely on syntactic processing and has previously been used as a non-syntactic cognitive task (Patston & Tippett, 2011). In Experiment 1, participants were asked to judge whether sound sequences were musical or environmental, and in Experiment 2, they were asked whether the sequences were basic or complex. The primary tasks in both experiments were the language comprehension and visuospatial search tasks, and the secondary tasks were the judgements about the sound sequences.

This manuscript was co-authored by myself, Glenn Schellenberg, Genevieve McArthur, and Bill Thompson. I contributed approximately 85% of the total work, including experimental design, stimuli creation, experiment programming, data collection, statistical analysis, and the preparation of the first draft of the manuscript. Glenn Schellenberg and Bill Thompson contributed to experimental design and result interpretation. Bill Thompson and Genevieve McArthur provided critical comments on the manuscript. Shayan Alam, a research student from the University of Toronto, Mississauga, tested approximately half of the participants in Experiment 1.

This chapter was prepared as:

Fiveash, A., Schellenberg, G., McArthur, G., & Thompson, W. F. (under preparation). Complexity or syntax? Music, language, and syntactic interference.

Chapter 4: Complexity

Complexity or Syntax? Music, Language, and Syntactic Interference

Anna Fiveash[1,3], Glenn Schellenberg[4], Genevieve McArthur[2,3], & William F. Thompson[1,3]

[1]Department of Psychology, Macquarie University

[2]Department of Cognitive Science, Macquarie University

[3]ARC Centre of Excellence in Cognition and its Disorders, Macquarie University

[4]Department of Psychology, University of Toronto, Mississauga

Chapter 4: Complexity

In music and language, discrete units of information are combined to create complex, meaningful sequences. Individual words are combined to form sentences that can be used to discuss complex concepts, such as atomic particles and the nature of the universe. Individual notes are combined into melodies and phrases that can arouse emotions, bring back memories, and unite people who might be unable to communicate any other way. The set of rules used to combine words or notes into meaningful sequences is called syntax. Syntactic sequences have hierarchical structure and contain elements that vary in importance within that structure. Furthermore, syntax describes how every element relates to every other element within a hierarchical network of interconnections. Syntax allows a listener to understand that a speaker is talking about *our place in the universe* and not *the universe's place in us*, and for a musician to elicit emotion in listeners by creating tension in music and then releasing it. These similarities provoke the question: are music and language, our two most syntactically complex systems, processed in a similar way?

One approach to this question is to examine whether interference effects are observed when complex music and language syntax are processed simultaneously. A syntactic interference effect would be expected if both domains were drawing upon the same limited-capacity syntactic resources (Patel, 2008). Previous research has revealed such interference effects when music syntax is processed at the same time as language syntax, suggesting the two domains engage a limited pool of shared syntactic processing resources (Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009; Fiveash & Pammer, 2014; Hoch, Poulin-Charronnat, & Tillmann, 2011; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Slevc, Rosenberg, & Patel, 2009). A common experimental design is to burden syntactic processing in music and language simultaneously. In such designs, stimuli include sung complex sentences with out-of-key notes on important words (Fedorenko et al., 2009), and out-of-key chords presented at the same time as grammatical errors (e.g., gender violations in German) in language (Koelsch et al., 2005).

However, it is difficult to infer from such studies whether interference arises exclusively from syntactic processes or from other sources such as distraction, error detection, or working memory costs that are introduced by syntactic errors or increased complexity in syntax (Fiebach, Schlesewsky, & Friederici, 2002; Fiveash, McArthur, & Thompson, under review; Tillmann & Bigand, 2015). Many researchers have attempted to control for the possibility that out-of-key elements are merely distracting. Such studies have included control conditions that involve a timbre change (Fiveash & Pammer, 2014; Slevc et al., 2009) or an increase in loudness (Fedorenko et al., 2009; Kunert, Willems, Casasanto, Patel, & Hagoort, 2015) instead of a syntactic violation, to try and match out-of-key elements for distraction. However, it has been suggested that such control stimuli may not compare to the level of sensory violation introduced by out-of-key elements, and are therefore not an adequate control (Tillmann & Bigand, 2015). It is therefore an open question as to whether syntactic interference effects are specific to syntax, or can be explained by more general processes.

Another experimental approach used to investigate syntactic interference effects between language and music is the combination of out-of-key notes or chords in music with syntactic or semantic manipulations in language. Studies using this approach tend to show an interference effect with syntactic manipulations, but not with semantic manipulations, suggesting that the effect is specific to syntax. For example, out-of-key (Slevc et al., 2009) and unexpected (Hoch et al., 2011) musical elements have been shown to increase reading times of syntactic garden path sentences and increase response times to syntactic violations in language in a lexical decision task (respectively). In contrast, no interference was shown for semantically unexpected words, suggesting a syntax-specific effect. However, a problem with this comparison is that unexpected semantic words are often more ambiguous, and hence less “surprising”, than syntactic errors in language (Tillmann & Bigand, 2015). Thus, interference may occur with syntactic errors but not semantic errors simply because syntactic errors are more salient and hence more distracting. This idea is supported by a study similar to Slevc et al. (2009) that used semantic garden path sentences instead of syntactic garden path sentences (Perruchet & Poulin-Charronnat, 2013). The results revealed that a music syntax violation resulted in longer reading times for semantic garden path sentences, but not sentences with a semantic error—the same pattern as the syntactic garden path sentences in Slevc et al. (2009). In sum, it is still not clear whether interference effects associated with syntactic violations reflect overlap in syntactic processes specifically, or are a consequence of overlap in more general processing resources related to complexity and error detection.

Given the difficulty in interpreting the effect of music syntax violations on language syntax processing, it would be valuable to demonstrate syntactic interference effects using stimuli that do not involve syntactic violations. One method of avoiding syntactic and semantic violations is to present sentences without violations that instead vary in syntactic complexity. For example, Fedorenko et al. (2009) and Kunert et al. (2015) manipulated complexity in language by presenting participants with two types of sentences: simple sentences that contained subject-extracted relative clauses (e.g., *the guest* that kissed the host *brought a cake to the party*), and more complex sentences that contained object-extracted relative clauses (e.g., *the guest* that the host kissed *brought a cake to the party*). The latter sentence is more difficult to process than the former because it involves a long distance dependency between the words *that* and *kissed*

(Fedorenko et al., 2009; Fiebach et al., 2002). Fedorenko et al. (2009) presented participants with subject-extracted and object-extracted relative clauses that were sung on melodies that contained no manipulation, an out-of-key note, or a note that increased in loudness. They found that out-of-key notes resulted in poorer comprehension of complex sentences, but not simple sentences, and that poorer comprehension did not occur when the same note was sung with an increase in loudness.

Kunert et al. (2015) conducted a similar study to Fedorenko et al. (2009) using functional magnetic resonance imaging (fMRI). Kunert et al. (2015) observed overlapping brain activation in Broca's area when complex sentences were sung with out-of-key notes. This effect did not occur for simple sentences or with a loudness increase, suggesting shared networks for processing music and language syntax. However, complex sentences also introduce increased working memory demands that are difficult to disentangle from syntactic effects. For example, Fiebach et al. (2002) conducted an event-related potential (ERP) study which showed that complex sentences resulted in both increased working memory demands, and increased syntactic integration demands, and that these two processes could be observed separately in the brain. As such, complexity manipulations in language stimuli could also be related to the increased processing cost involved in holding words in working memory before the sentence can be parsed.

A number of general cognitive processes, including working memory, have been suggested to make up a domain-general cognitive control system referred to as the *multiple-demand system* (Duncan, 2010; Fedorenko, 2014; Fedorenko, Duncan, & Kanwisher, 2013). The multiple-demand system has been linked to domain-general cognitive control and executive functions, and is activated with cognitive challenges involving problem solving, focused attention, and novel situations (among others) across a number of different task types (Duncan, 2010; Fedorenko et al., 2013).

Fedorenko, Behr, and Kanwisher (2011) have suggested that the language network is distinct from the multiple-demand system and uses largely separate processing resources. They further suggested that music does not share neural circuitry with the language network, and instead, draws upon the multiple-demand system. To test this claim, Fedorenko et al. (2011) conducted an fMRI study with the aim of isolating the so-called language network, and then comparing this network with brain activation from other stimuli such as music, mathematics, spatial working memory, verbal working memory, and interference tasks. These tasks were designed to activate arithmetic, working memory, and cognitive control aspects of the multiple-demand system. To isolate the language network, Fedorenko et al. (2011) presented participants in an fMRI scanner with both non-words and sentences, and subtracted brain activation to non-words from brain activation to sentences. To isolate the multiple-demand system, they presented participants with hard (complex) and easy (simple) tasks tapping into working memory, cognitive control and arithmetic. They then subtracted brain activation from the simple tasks from brain activation to the complex tasks to isolate processing related to these individual tasks. To isolate music processing—which was hypothesised to use the multiple-demand system—participants were presented with intact and scrambled music, and brain activation to scrambled music was subtracted from brain activation to intact music.

These comparisons revealed that the language network was not activated by non-language tasks, suggesting a distinction between the language network and the multiple-demand system. This finding led the authors to suggest that the multiple-demand system (and the processing of music) is functionally distinct from the language network. However, they do suggest that the language network draws on multiple-demand resources, and connections between the language network and the multiple-demand system are further discussed in Fedorenko (2014). In contradiction to their claim that language and music were processed with distinct resources, Fedorenko et al. (2011) mentioned that the music comparison activated a number of areas in the language network. This overlap was not significant when multiple comparisons were corrected for; however, it is possible that the particular comparison they used (intact compared to scrambled music) was not appropriate to isolate music processing. Further, a commentary by Caplan (2014) outlined a number of reasons why the specific contrasts and language localising tasks, combined with the temporal insensitivity of fMRI, might not be appropriate to draw the conclusion that the multiple-demand system and the language network are distinct.

In addition to the concerns raised by Caplan (2014), it is also possible that the comparison between intact and scrambled music was not appropriate to draw the conclusion that music processing is distinct from the language network. The subtraction of brain activation to scrambled music from brain activation to intact music is not comparable to the subtraction of simple tasks from complex tasks. In addition, the stimuli in the intact versus scrambled conditions differed in a number of acoustic features. First, scrambled music disrupts rhythmic and melodic expectations, and is far from a "simple" version of intact music. To the contrary, scrambled music may have increased the perceived complexity of the music. Eerola (2016) suggested that complexity in music relates to how unexpected elements in the music are. Therefore, scrambling music is likely to result in more complex stimuli than intact songs that conform to expectations. Second, the intact music condition consisted of pop and rock songs from the 1950's and 1960's. To scramble this music, Fedorenko et al. (2011) first converted it to musical instrument digital interface (MIDI) representations in which performance expression was removed entirely, and then manipulated pitch and timing information within these sequences. The vast acoustic differences between the intact and scrambled music, and the theoretical reasons why scrambled music should not be considered an “easy” version of intact music (Eerola, 2016), suggest that the subtraction of scrambled music from intact music may not reflect a pure measure of music processing. Therefore, conclusions about the specificity of language processing in relation to music processing seem premature. It is unclear why the authors did not use a complexity manipulation for music (e.g., simple compared to complex music), as this manipulation would have been more comparable to the simple and complex tasks used for the other stimuli. Based on the considerations presented above, it is clear that there is a need for music and language stimuli that involve careful manipulations of complexity, and comparable control tasks.

To this end, the current experiments manipulated complexity in music (syntactic) and environmental (non-syntactic) sound sequences without introducing syntactic violations. Music and environmental sound sequences were presented with a concurrent language comprehension task (syntactic) or a concurrent visuospatial search task (non-syntactic). To manipulate complexity, basic and complex versions of both music and environmental stimuli were created. Basic stimuli consisted of two repeating elements (chords or environmental sounds). Complex stimuli consisted of a sequence of multiple different elements (chords or environmental sounds). Increasing complexity in music introduces a greater number of tonal dependencies between elements, and hence results in a more complex syntactic structure compared to basic sequences of two alternating chords (Lerdahl & Jackendoff, 1983). Increasing complexity in environmental sounds does not introduce functional dependencies between elements or syntax, as environmental sounds are not structured hierarchically. Greater interference was therefore expected from complex music than from complex environmental sounds on a language comprehension task, whereas minimal interference from the basic stimuli was expected. The visuospatial search task was included as a control condition to measure general levels of distraction introduced by each auditory condition when the task was not syntactic.

Increasing the complexity of either music or environmental sound sequences should increase the burden on working memory. However, whereas complex music should engage *syntactic* processing, complex environmental sound sequences should not. To test this hypothesis, we measured working memory capacity (WMC) and included it as a covariate throughout our analyses. If observed syntactic interference effects disappear when WMC is statistically controlled for, then this outcome could indicate that working memory can explain a large amount of the variance related to syntactic processing. However, if syntactic interference effects are still observed when WMC is controlled for, then this outcome would suggest that working memory cannot fully account for syntactic processing. Musical training was also measured, as previous research has suggested that syntactic interference is stronger for musicians (Patston & Tippett, 2011).

Experiment 1

Method

Participants. Fifty first-year undergraduate students from the University of Toronto, Mississauga participated in this study for course credit (M_{age} = 18.5, range: 17-27, 45 females). Of these participants, 36 spoke English as their first language, 14 people spoke English as their second language, and all except one learnt English before the age of five (range: 2-5 years of age). One participant learnt English at age seven, but did not perform differently to the rest of the group so was included in the analyses. Second languages varied widely (Asian, Middle Eastern, and European languages), reflecting the population of Mississauga. Thirty-five participants reported speaking a second language fluently, and 15 reported speaking only English. Eight participants reported five or more years of private music lessons, and 10 reported currently being musically active. Forty-four participants reported listening to music regularly (M_{mins} = 172, *SD* = 140). No participants reported having absolute pitch.

Design. Experiment 1 was a 2 (complexity: basic, complex) by 2 (auditory stimulus: music, environmental sounds) by 2 (task: language comprehension, visuospatial search) repeated measures design. There were four auditory stimulus conditions: basic music, basic environmental, complex music, and complex environmental. While listening to one of the auditory stimulus conditions, participants completed either a language comprehension task or a visuospatial search task. The language comprehension task involved reading a complex sentence and then answering a comprehension question. The visuospatial search task involved detecting a difference between two images. Language comprehension and visuospatial search trials were presented in separate blocks that were counterbalanced across participants. There were 20 different stimuli in each auditory stimulus condition (i.e., 80 stimuli) that were randomly paired for each participant with 80 complex sentences and 80 visuospatial search images. There were 160 trials in total. The experiment was programmed in Matlab (version 2016b; Mathworks) and presented with Psychtoolbox (Brainard, 1997).

Auditory stimuli (music, environmental sounds). The *complex music* (syntactic) condition consisted of MIDI chord sequences adapted from Slevc et al. (2009). These sequences were all in the key of C major, were tonal, and ended on perfect authentic cadences. Only chord sequences with eight or nine chords were used.

As there were not enough chord sequences with eight or nine chords, some of the chord sequences were reversed to create 20 novel sequences. For the reversed sequences, the last two chords were always retained in the final position. Some of the chord sequences were transposed to introduce variation across stimuli in pitch height. Thus, sequences were presented in the major keys of C (n = 4), C$^{\#}$ (n = 4), D (n = 3), D$^{\#}$ (n = 3), E (n = 3), and F (n = 3). All stimuli played for exactly five seconds (to correspond with task presentation). Therefore, chord sequences with eight chords were played at 96 beats per minute (bpm), and chord sequences with nine chords were played at 108bpm. All music stimuli were imported into GarageBand, and played with the MIDI instrument Steinway grand piano. An initial pilot study (n = 16) failed to reveal any differences between groups across any of the conditions. Therefore, to increase syntactic processing, a modulation was included in each chord sequence. The modulation was always introduced on the fifth chord, and involved a semitone shift up in key (e.g., a sequence in C would shift to C$^{\#}$). All subsequent chords were compatible with the new key, and the sequences all ended with a perfect authentic cadence in the new key. These measures ensured that the manipulation would be perceived as a modulation and not an out-of-key chord. Modulations require listeners to integrate the new key into the current musical context, which should place high demands on syntactic resources (Patel, 2008). The *basic music* condition consisted of the last two chords in each chord sequence repeated four times (e.g., eight chords in total). All basic stimuli were played at 96bpm.

Complex environmental stimuli were created by combining one-second environmental sounds from the norms for environmental sound stimuli (NESSTI) database (Hocking, Dzafic, Kazovsky, & Copland, 2013). This database contains 110 different environmental sounds from multiple categories. We excluded human sounds, musical sounds, and any sounds that had a noticeably lower intensity and reduced sound spectrum compared to the other recordings (e.g., a moving bicycle and cutting scissors). This left 69 environmental sounds. All sounds were normalised so their intensity level was comparable to the chord sequences. Thirty environmental sounds were then randomly selected and allocated to one of the 30 different chords in the complex music stimuli. For example, the C major chord always corresponded to an elephant noise. A chord sequence of: C, F, C, Am, F, Em, Dm, G7, C had an environmental sound analogue of: elephant, sheep, elephant, owl, sheep, bee, lion, canary, elephant. This mapping of environmental sounds to chords resulted in a pattern within the environmental sequences that was comparable to the chord sequences. However, unlike chord sequences, sequences of environmental sounds were not syntactic, as there were no dependencies between elements. Thus, each chord sequence in the complex music condition had a non-syntactic environmental sound analogue.

Environmental sound sequences were created using Audacity. To create 5 second sequences, the “change tempo” effect in Audacity was applied, which allowed the duration of stimuli to be adjusted without changing pitch. The *basic environmental* condition was created in a similar way to the basic music condition, by repeating the final two environmental sounds from each sequence four times (i.e., eight environmental sounds in total). The change tempo effect in Audacity was then applied to make the sequences 5 seconds long.

Task type (language comprehension, visuospatial search). In the language comprehension task participants were presented with *complex sentences* in the form of object-extracted relative clauses used by Fedorenko et al. (2009) and Staub (2010). Sentences averaged 12 words (range: 10-15), and each had an accompanying comprehension question. The 36 sentences from Fedorenko et al. (2009) had comprehension questions already designed. For the extra 44 sentences, comprehension questions were created to parallel the Fedorenko et al. (2009) questions. As an example taken from Fedorenko et al. (2009), participants were presented the sentence: *The cop who the spy met wrote a book about the case*, and then given the comprehension question: *Did the cop write a book about the case?* In that example, the correct answer is yes. Across all stimuli, the correct answer was yes for half the stimuli, and no for the other half.

The *visuospatial search task* was designed by Patston and Tippett (2011), and adapted to be presented on the computer. One trial consisted of two square designs side-by-side, and there was always one small difference between the two designs to be detected. The designs were made up of 12 different geometric shapes and a number of red, blue, green, and yellow coloured dots (see Figure 1). The design was separated into four quadrants. Within each pair of designs, a dot would either change colour or move position within a quadrant. The task was designed to require a visual search for the difference rather than it being immediately obvious.

Image removed due to copyright, please see citation below

Figure 1. An example of a visuospatial search trial, designed by Patston and Tippett (2011). The yellow circle in the bottom right quadrant is in a different position in each design. Circles could also change colour instead of position.

Procedure. Participants first signed the information and consent form, and filled in a musical experience and demographic questionnaire. Participants heard examples of the four types of auditory stimuli (basic and complex music or environmental sounds), and were told that in the language trials they would be reading sentences and answering comprehension questions. In the visuospatial search trials, they were told to indicate the quadrant that contained the dot with the different colour or position (Patston & Tippett, 2011). Participants had two practices with the language comprehension task and the visuospatial search task while listening to the different types of auditory stimuli. After each practice trial, participants were shown the correct answers. Once it was verbally confirmed that participants understood the task, the experiment proper began.

Language and visuospatial search trials were blocked and counterbalanced so that participants first performed the language comprehension task or the visuospatial search task. There were three breaks in each block, and a break between blocks. For both tasks, a fixation-cross first appeared on the screen for 1 second. A sentence or a visuospatial search trial then appeared on the screen, and was accompanied by an auditory stimulus presented concurrently for 5 seconds. In the language task, there was a 1 second pause after the paired stimuli ended. A comprehension question then appeared on the screen. To minimise ceiling effects and use of memory to answer the comprehension question, the same auditory stimulus played again and participants only had 6 seconds to answer the question before the next trial began. Participants answered the comprehension question using the keyboard, by indicating yes (press z) or no (press m) in response to the question. In the visuospatial task, participants indicated which quadrant contained the difference (1, 2, 3, or 4). The music did not play again in this time, and participants were not limited in response time. After participants provided the relevant response, they were asked to indicate whether the auditory stimulus was musical or environmental. This question was included to check that participants were not ignoring the auditory stimulus. See Figure 2.

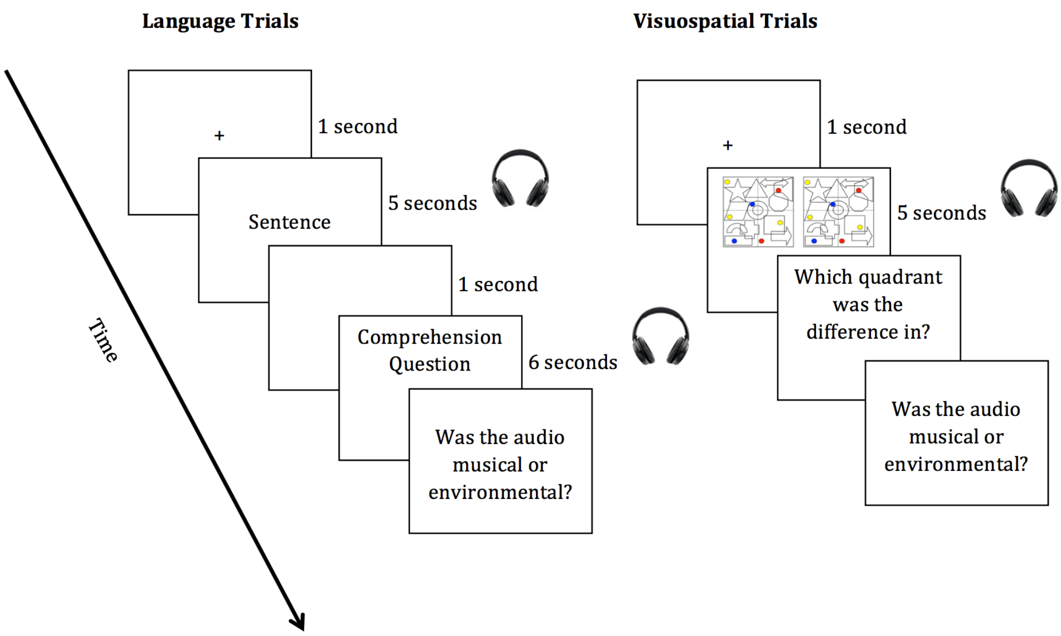

Figure 2. Timeline of stimulus presentation for the language and visuospatial search trials. Headphones represent auditory stimulus presentation.

At the end of the experiment, participants completed the letter-number sequencing subtest of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 2008). This test is a measure of WMC in which participants are presented with a sequence of numbers and letters and asked to rearrange them so that all the numbers occur first in ascending order, followed by the letters in alphabetical order. The test was adapted to be presented on the computer, and participants wrote their answers on a sheet of paper. They were given instructions and five practice trials consisting of two- and three-element sequences. Participants were asked to indicate their answer in all practice trials, which allowed an assessment of whether they understood the task. Once the practice trials were completed and it was clear the participant understood the task, the main test commenced. There were three trials for each sequence length (2-8), resulting in 21 trials for a total score out of 21. The test took approximately 10 minutes, and the whole testing session took approximately 50 minutes.

Scoring. To measure performance on the language task, d prime (d’) values were calculated based on signal detection theory (Stanislaw & Todorov, 1999). D prime is a measure of sensitivity to the signal based on hits and false alarm rates, where the z score for false alarms is subtracted from the z score for hits. Extreme values (e.g., 1 or 0) were corrected for: scores of 1 were replaced by (n – 0.5) and scores of 0 were replaced by 0.5/n, where n = the number of signal or noise trials. See Stanislaw and Todorov (1999) for more detail.

Because there were four options to choose from in the visuospatial search task, accuracy of response was measured for the analysis. Therefore, for each auditory stimulus condition, participants were deemed to have made a correct response if they correctly identified the quadrant in which the difference occurred, and incorrect if they identified any other quadrant. This scoring procedure provided accuracy scores for each participant in each condition. Accuracy ratings for whether participants judged the auditory stimulus as musical or environmental were also analysed as proportion correct.

Reaction times were recorded for each trial for the language and visuospatial search tasks, and average RTs were calculated for each condition for each participant. As participants were not limited in their response time for the visuospatial search task, RT scores that were more than three standard deviations above the average RT for each participant were excluded from the analysis. Language comprehension RTs were all under 6 seconds.

Analysis. Because the language comprehension and visuospatial search trials were measured on different and unrelated scales, they were analysed separately. Repeated measures analyses of variance (ANOVAs) were performed separately on the language comprehension d' scores and the visuospatial search accuracy scores, with the factors of complexity (basic, complex) and auditory stimulus (music, environmental sounds). To investigate the role of WMC on performance, and whether this differed depending on auditory stimulus, these ANOVAs were also performed with the added covariate of WMC (as measured in the letter-number sequencing sub-test). All results of ANOVAs are reported first without covariates, and then with covariates, as suggested by Simmons, Nelson, and Simonsohn (2011). Significant between-subjects effects of WMC were followed up with bivariate correlations between WMC and performance on each task. The same analyses were repeated for the RT data. A preliminary analysis including years of private musical training as a covariate in the ANOVAs for both the language comprehension and visuospatial search results revealed no differences and no between-subjects effects of musical training, so all participants were analysed as one group.

Results

Language task. A repeated measures ANOVA was conducted on the language comprehension d' scores with the factors complexity (basic, complex) and auditory stimulus type (music, environmental sounds). There was no main effect of complexity, $F(1, 49) = .05$, $p = .83$, no main effect of auditory stimulus type, $F(1, 49) = 1.06$, $p = .31$, and no interaction, $F(1, 49) = .33$, $p = .57$. These results suggest that background auditory stimulus (basic or complex music or environmental sounds) did not affect language comprehension. When WMC was included as a covariate, there were still no main effects or interaction effects.[1] However, there was a significant between-subjects effect of WMC, $F(1, 47) = 5.96$, $p = .02$, $\eta^2 = .113$.

To investigate the direction of this between-subjects effect, d' scores were calculated for language comprehension averaged across all of the auditory stimulus conditions (basic music, basic environmental, complex music, and complex environmental). Bivariate correlations revealed a positive correlation between WMC and language comprehension performance in general ($r = .34$, $p = .018$), suggesting that participants with higher WMC performed better on the language comprehension task. This result would be expected considering participants had to hold the sentence in memory to answer the comprehension question. However, WMC did not affect language comprehension differently depending on auditory stimulus type.

To investigate whether there were differences in RT for language comprehension judgements depending on auditory stimulus condition, a 2 (complexity: basic, complex) by 2 (auditory stimulus: music, environmental sounds) repeated measures ANOVA was conducted. There was no main effect of complexity, $F(1, 49) = .26$, $p = .61$, no main effect of auditory stimulus, $F(1, 49) = .06$, $p = .81$, and no interaction, $F(1, 49) = .10$, $p = .75$, suggesting that background auditory stimulus did not affect RTs for language comprehension judgements. With WMC added as a covariate, there were still no main effects or interaction effects.[2] In contrast to the d'

[1] There was no main effect of complexity, $F(1, 47) = .01$, $p = .92$, no main effect of auditory stimulus type, $F(1, 47) < .001$, $p = .99$, no interaction between complexity and auditory stimulus type, $F(1, 47) = .52$, $p = .47$, no interaction between complexity and WMC, $F(1, 47) = .03$, $p = .87$, no interaction between auditory stimulus type and WMC, $F(1, 47) = .03$, $p = .86$, and no three-way interaction, $F(1, 47) = .45$, $p = .51$.

[2] There was no main effect of complexity, $F(1, 47) = .02$, $p = .88$, no main effect of auditory stimulus type, $F(1, 47) = .18$, $p = .67$, no interaction between complexity and auditory stimulus type, $F(1, 47) = .29$, $p = .60$, no interaction between complexity and WMC, $F(1, 47) = .01$, $p = .93$, no interaction between auditory stimulus type and WMC, $F(1, 47) = .12$, $p = .73$, and no three-way interaction, $F(1, 47) = .45$, $p = .51$.

scores, there was no between-subjects effect of WMC, $F(1, 47) = .94$, $p = .34$. These results suggest that WMC did not affect RT for language comprehension judgements.

Visuospatial search task. A repeated measures ANOVA with the same factors of complexity (basic, complex), and auditory stimulus (music, environmental sounds) was conducted on the accuracy scores for the visuospatial search task. There was no main effect of complexity, $F(1, 49) = .002$, $p = .97$, no main effect of auditory stimulus, $F(1, 49) = 1.16$, $p = .29$, and no interaction effect, $F(1, 49) = 1.20$, $p = .28$. Similarly to the language task, when WMC was added a covariate, there were still no main effects or interaction effects,[3] but there was a significant between-subjects effect of WMC, $F(1, 47) = 8.48$, $p = .005$, $\eta^2 = .15$. To explore the direction of this between-subjects effect, the average visuospatial search accuracy was calculated across all four conditions for each participant. Bivariate correlations revealed a significant positive correlation between WMC and visuospatial search accuracy ($r = .39$, $p = .005$), suggesting that people with higher WMC performed better on the visuospatial search task. This result would be expected, considering links between WMC and processing speed (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002).

To investigate whether RTs on the visuospatial search task differed depending on auditory stimulus condition, we conducted a 2 (complexity: basic, complex) by 2 (auditory stimulus: music, environmental sounds) repeated measures ANOVA on the RT data. This analysis also showed no main effect of complexity, $F(1, 49) = .11$, $p = .74$, no main effect of auditory stimulus, $F(1, 49) = 1.2$, $p = 28$, and no interaction, $F(1, 49) = 3.52$, $p = .07$. The same ANOVA with an added covariate of WMC showed no

[3] There was no main effect of complexity, $F(1, 47) = .35$, $p = .56$, no main effect of auditory stimulus type, $F(1, 47) = .94$, $p = .34$, no interaction between complexity and auditory stimulus type, $F(1, 47) = .02$, $p = .90$, no interaction between complexity and WMC, $F(1, 47) = .42$, $p = .52$, no interaction between auditory stimulus type and WMC, $F(1, 47) = .58$, $p = .45$, and no three-way interaction, $F(1, 47) = .21$, $p = .65$.

main effects or interaction effects,[4] and no between-subjects effect of WMC, $F(1, 49) = 2.3$, $p = .13$, suggesting that WMC did not affect RTs on the visuospatial search task.

Music or environmental sound judgement task. The judgements on whether auditory stimuli were musical or environmental were very high across all auditory stimulus conditions while performing both the language comprehension and visuospatial search tasks. When concurrently completing the language comprehension trials, participants showed ceiling effects for judgements of whether the auditory stimulus was musical or environmental for all conditions: basic environmental ($M_{\text{accuracy}} = .97$, $SD = .06$), basic music ($M_{\text{accuracy}} = .97$, $SD = .06$), complex environmental ($M_{\text{accuracy}} = .97$, $SD = .04$), complex music ($M_{\text{accuracy}} = .98$, $SD = .05$). While completing the visuospatial search task, performance was similarly high: basic environmental ($M_{\text{accuracy}} = .98$, $SD = .07$), basic music ($M_{\text{accuracy}} = .97$, $SD = .05$), complex environmental ($M_{\text{accuracy}} = .98$, $SD = .05$), complex music ($M_{\text{accuracy}} = .98$, $SD = .04$). These results suggest that participants successfully processed the auditory stimuli to some level.

Discussion

Experiment 1 revealed no significant effects of auditory stimulus (music or environmental sounds) or complexity (basic or complex) on language comprehension or visuospatial search. However, WMC was positively correlated with performance on the language comprehension task and the visuospatial search task, showing that participants with higher WMC performed better on both tasks, but that this effect did not differ depending on concurrent auditory stimulus condition. The absence of any difference in

---

[4] There was no main effect of complexity, $F(1, 47) = .06$, $p = .81$, no main effect of auditory stimulus type, $F(1, 47) = .33$, $p = .57$, no interaction between complexity and auditory stimulus type, $F(1, 47) = .18$, $p = .67$, no interaction between complexity and WMC, $F(1, 47) = .02$, $p = .88$, no interaction between auditory stimulus type and WMC, $F(1, 47) = .13$, $p = .72$, and no three-way interaction, $F(1, 47) = .01$, $p = .91$.

language comprehension or visuospatial search depending on auditory stimulus or complexity was surprising, considering the substantial differences between auditory stimuli.

We suggest three possible reasons why syntactic interference was not observed. First, participants may have primarily ignored the auditory stimuli. Categorising auditory stimuli into musical or environmental is a very simple and easy judgement, as evidenced by the ceiling effects in the data. Judging whether a sequence is musical or environmental does not rely on working memory processes and only requires participants to pay attention to the auditory stimulus very briefly to make the judgement. Although research has suggested that the processing of syntax is largely automatic and occurs without explicit attention (e.g., while watching silent movies; Koelsch & Jentschke, 2008), attention has been shown to affect the strength of syntactic processing (Loui, Grent-'t-Jong, Torpey, & Woldorff, 2005). For these reasons, it is possible that participants were not engaging with the auditory stimuli to the extent that interference could be observed. Second, it is possible that the music stimuli were not complex enough for syntactic interference to occur. To elicit interference with language comprehension, Fedorenko et al. (2009) had to speed up their stimuli and include out-of-key elements. Third, it is possible that the large variety of languages our sample of participants spoke along with English might have contributed to our null findings on the language comprehension task. The sentences we used had a very particular structure (object-extracted relative clauses) that may have been less familiar to some participants. Experiment 2 was designed to address these issues.

Experiment 2

The aim of Experiment 2 was to address the limitations of Experiment 1 in three ways. First, participants were asked to make a more difficult judgement—whether the auditory stimulus was basic or complex. To judge the complexity of a stimulus, participants would have to pay closer attention to the sound sequence, and engage working memory. In a pilot test, participants were asked to process the stimuli at an even deeper level (e.g., whether the chord sequence changed key, or whether the environmental sounds changed category between animal and other). The pilot test revealed that this manipulation resulted in floor effects on all tasks, suggesting that participants were unable to prioritise two tasks simultaneously. Complexity judgements were therefore used. Second, to make the music more complex, an additional key modulation was added to the complex music stimuli in Experiment 2. Modulations require the new key to be integrated into the current key, and therefore two modulations should tax syntactic processing more than only one modulation within a sequence. Third, to ensure the null findings were not because of our sample, we tested this study in Australia with participants who spoke English as their first language, and were primarily monolingual.

Method

Participants. Thirty-seven participants from Macquarie University in Australia participated in this study for course credit or $15 ($M_{age}$ = 25.6 years, SD = 9.3, range: 18-53; 26 females, 10 males, and one who did not identify either way). All participants spoke English as their first language, with the exception of one participant who was raised bilingual (English and Cantonese). Twenty-nine participants reported speaking only English fluently, and 32 reported being born in an English-speaking country. None reported any language disorders, and only one person reported a slight hearing decrement. Eighteen participants indicated they had five or more years of private music training, and 14 reported being currently musically active. All participants reported listening to music, with an average listening time of 143 minutes per day (*SD* = 112 mins). No participants reported having absolute pitch.

Design and procedure. Experiment 2 was the same design as Experiment 1, with a 2 (complexity: basic, complex) by 2 (auditory stimulus: music, environmental sounds) by 2 (task: language comprehension, visuospatial search) repeated measures design. The language comprehension and visuospatial search tasks were identical to Experiment 1, as were the basic music, basic environmental, and complex environmental auditory stimulus conditions. To increase syntactic processing difficulty in the complex music sequences, we added an extra modulation into the chord sequences. The procedure was identical to Experiment 1, except instead of asking participants to indicate whether the stimuli were musical or environmental, participants were asked to indicate whether the sequences were basic or complex.

Complex music stimuli. For the new complex music sequences, the first modulation occurred on the fifth note (as in Experiment 1) and the second modulation occurred on either the seventh or eighth note, depending on the length of the sequence. All sequences ended in a perfect authentic cadence in the final key. Each modulation shifted the chords up in pitch by one semitone. For example, a sequence starting in the key of C shifted to the key of $C^{\#}$, then to the key of D, and resolved in the key of D. The first modulation (to $C^{\#}$) could be considered a transient modulation, and then the second modulation (to D) could be considered a direct modulation, as it resolved in a perfect authentic cadence to reinforce the new key. Although there were momentary out-of-key elements in the sequences, the sequences all resolved in the new key, suggesting a tonal shift, and engaging syntactic processing resources.

Scoring and analysis. As in Experiment 1, d' scores were calculated for the language comprehension task, and proportion correct was calculated for the visuospatial search task. Reaction times across all conditions for both tasks were also calculated as in Experiment 1. In addition to the primary task, participants also made auditory complexity judgements where they indicated whether each auditory stimulus was basic or complex. To analyse these data, d' scores were calculated where complex sounds were treated as the signal, and basic sounds were treated as the noise. Thus, hits corresponded to the correct identification of complex sequences, and false alarms corresponded to the incorrect identification of basic sequences (e.g., indicating the sequence was complex when it was basic). The d' scores therefore provided a measure of sensitivity to complexity for both music and environmental sound stimuli when participants were engaged in either the language comprehension task or the visuospatial search task.

The main analysis was the same as Experiment 1—ANOVAs were performed on the language comprehension d' scores, visuospatial search accuracy, and RT data for both tasks separately. Working memory capacity was subsequently included as a covariate in each ANOVA. Any significant main effects were followed up with paired-samples *t*-tests, and between-subject effects of WMC were investigated with bivariate correlations. Auditory complexity judgements (d' scores) were analysed with a 2 (task: language comprehension, visuospatial search) by 2 (auditory stimulus: music, environmental sounds) repeated measures ANOVA. Main effects were investigated with paired-samples *t*-tests. A final ANOVA was then conducted with WMC as a covariate in the ANOVA testing complexity judgements. All results of ANOVAs are reported first without covariates, and then with covariates, as suggested by Simmons et al. (2011). All Cohen's *d* values reported are based on repeated measures calculations, which take into account paired-samples correlations between variables. A preliminary analysis including years of private musical training as a covariate in the ANOVAs for both the language comprehension and visuospatial search results revealed no differences and no between-subjects effects of musical training, so all participants were analysed as one group.

Results

Language comprehension. To examine the effects of complexity (basic, complex) and auditory stimulus (music, environmental sounds) on the language comprehension d' scores, a repeated measures ANOVA was conducted. As in Experiment 1, the results of Experiment 2 provided no evidence for interference effects between music and language comprehension. There was no main effect of complexity, $F(1, 36) = .06$, $p = .81$, no main effect of auditory stimulus, $F(1, 36) = .91$, $p = .35$, and no interaction effect, $F(1, 36) = .06$, $p = .80$. The same analysis with the covariate of WMC showed no main effects or interaction effects,[5] and no between-subjects effect of WMC, $F(1, 35) = .38$, $p = .54$.

An ANOVA with the RT data also showed no main effect of complexity, $F(1, 36) = .17$, $p = .69$, no main effect of auditory stimulus, $F(1, 36) = 1.36$, $p = .25$, and no interaction, $F(1, 36) = 1.5$, $p = .23$. Including WMC as a covariate in the above analysis did not reveal any main effects or interaction effects,[6] but did reveal a significant between-subjects effect of WMC, $F(1, 35) = 7.3$, $p = .011$, $\eta^2 = .17$. Further

[5] There was no main effect of complexity, $F(1, 35) = .001$, $p = .98$, no main effect of auditory stimulus type, $F(1, 35) = 3.09$, $p = .09$, no interaction between complexity and auditory stimulus type, $F(1, 35) = .14$, $p = .71$, no interaction between complexity and WMC, $F(1, 35) = .01$, $p = .92$, no interaction between auditory stimulus type and WMC, $F(1, 35) = 2.4$, $p = .13$, and no three-way interaction, $F(1, 35) = .21$, $p = .65$.

[6] There was no main effect of complexity, $F(1, 35) = .48$, $p = .49$, no main effect of auditory stimulus type, $F(1, 35) < .001$, $p = .99$, no interaction between complexity and auditory stimulus type, $F(1, 35) = .60$, $p = .45$, no interaction between complexity and WMC, $F(1, 35) = .37$, $p = .55$, no interaction between auditory stimulus type and WMC, $F(1, 35) = .12$, $p = .74$, and no three-way interaction, $F(1, 35) = 1.32$, $p = .26$.

correlational analyses revealed that WMC was negatively correlated with language comprehension RTs averaged across all conditions (r = -.42, p = .01), suggesting that participants with higher WMC responded more quickly on the language comprehension questions in general.

Visuospatial search task. A repeated measures ANOVA on visuospatial search accuracy also showed no main effect of complexity, $F(1, 36) = .10$, $p = .75$, auditory stimulus, $F(1, 36) = 1.35$, $p = .25$, or interaction effect, $F(1, 36) = 2.71$, $p = 1.1$. When WMC was added as a covariate, there were still no main effects,[7] and no between-subjects effect of WMC, $F(1, 35) = 2.42$, $p = .13$. However, there was a marginally significant auditory stimulus by complexity interaction, $F(1, 35) = 4.54$, $p = .04$, $\eta^2 = .12$.

A repeated measures ANOVA on the RTs for the visuospatial search task with the factors of complexity (basic, complex) and auditory stimulus (music, environmental sounds) revealed a main effect of complexity, $F(1, 36) = 13.8$, $p = .001$, $\eta^2 = .28$, no main effect of auditory stimulus, $F(1, 36) = 1.06$, $p = .31$, and no interaction effect, $F(1, 36) = 2.33$, $p = .14$. Paired-samples t-tests with Holm-Bonferroni corrected p values for six comparisons revealed that RTs for the visuospatial search judgements were significantly slower in the complex environmental condition ($M_{\text{seconds}} = 1.02$, $SD = .33$) compared to the basic environmental condition ($M_{\text{seconds}} = .93$, $SD = .34$), $t(36) = 3.9$, $p' < .001$, $d = .63$. Reaction times were also significantly slower in the complex environmental condition compared to the basic music condition ($M_{\text{seconds}} = .94$, $SD = .31$), $t(36) = 3.8$, $p' = .005$, $d = .61$. No other paired comparisons yielded significant results: basic environmental and basic music, $t(36) = .32$, $p' = .75$, basic environmental

[7] There was no main effect of complexity, $F(1, 35) < .001$, $p = .98$ and no main effect of auditory stimulus type, $F(1, 35) = .47$, $p = .50$. There was no interaction between complexity and WMC, $F(1, 35) = .004$, $p = .95$, no interaction between auditory stimulus type and WMC, $F(1, 35) = .15$, $p = .70$, and no three-way interaction, $F(1, 35) = 3.02$, $p = .09$.

and complex music ($M_{seconds}$ = .98, *SD* = .31), $t(36) = 1.59$, *p'* = .36, basic music and complex music, $t(36) = 1.37$, *p'* = .36, and complex music and complex environmental, $t(36) = 1.87$, *p'* = .28. This pattern of results suggest that the complex environmental condition was driving the main effect of complexity, with longer RTs than all other conditions except the complex music condition.

When the above ANOVA was performed again with WMC as a covariate, the main effect of complexity was non-significant, $F(1, 35) = 1.35$, $p = .25$, and there was a significant between-subjects effect of WMC, $F(1, 35) = 5.57$, $p = .02$, $\eta^2 = .14$. All other main effects and interaction effects were non-significant.[8] Bivariate correlations showed that RTs averaged across all conditions for the visuospatial search task were significantly negatively correlated with WMC ($r = -.37$, $p = .024$), suggesting that visuospatial search judgements were performed more quickly by participants with higher WMC than by those with lower WMC.

Complexity judgements. To evaluate whether the type of task participants were engaged in had an effect on complexity judgements (judging whether the auditory stimulus was basic or complex), a 2 (task: language comprehension, visuospatial search) by 2 (auditory stimulus: music, environmental sounds) repeated measures ANOVA was conducted on the d' values for complexity judgements. This analysis revealed a main effect of task, $F(1, 36) = 15.26$, $p < .001$, $\eta^2 = .30$, a main effect of auditory stimulus, $F(1, 36) = 4.91$, $p = .03$, $\eta^2 = .12$, but a non-significant interaction, $F(1, 36) = 3.81$, $p = .059$. Paired-samples *t*-tests were performed and the Holm-Bonferroni correction was applied for six multiple comparisons (adjusted *p* values reported). There was no difference between the environmental ($M = 2.55$, $SD = 1.04$)

[8] There was no main effect of auditory stimulus type, $F(1, 35) = .14$, $p = .71$, no interaction between complexity and auditory stimulus type, $F(1, 35) = 1.4$, $p = .24$, no interaction between complexity and WMC, $F(1, 35) = .03$, $p = .86$, no interaction between auditory stimulus type and WMC, $F(1, 35) = .46$, $p = .50$, and no three-way interaction, $F(1, 35) = 2.79$, $p = .10$.

and musical (M = 2.67, SD = 1.30) complexity judgements when participants were engaged in the language task, $t(36) = .73$, $p' = .74$. However, there was a significant difference between the environmental (M = 2.84, SD = .96) and musical (M = 3.24, SD = .86) complexity judgements when participants were engaged in the visuospatial search task, $t(36) = 4.23$, $p' < .001$, $d = .71$. See Figure 3. These results suggest that participants were significantly better at judging complexity in music than environmental sounds when they were concurrently performing a visuospatial search task, but that this pattern did not occur while they were engaged in a language comprehension task.

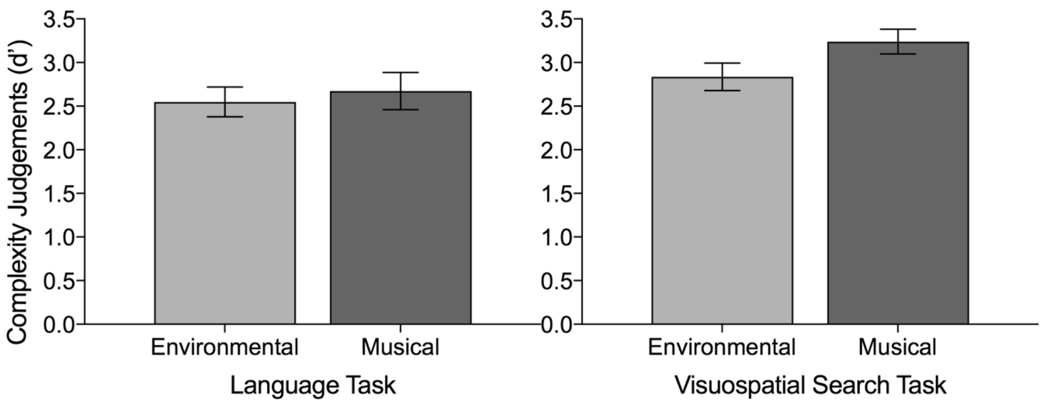

Figure 3. Complexity judgements for environmental and musical stimuli while participants performed either a language comprehension task or a visuospatial search task. Error bars indicate one standard error either side of the mean.

A closer examination revealed that musical complexity judgements were significantly better when participants were concurrently engaged in the visuospatial search task compared to when they were concurrently engaged in the language comprehension task, $t(36) = 4.01$, $p' < .001$, $d = .63$. However, there was no difference (after correcting for multiple comparisons) between environmental complexity judgements when participants were completing the visuospatial search task compared to when they were completing the language comprehension task, $t(36) = 2.4$, $p' = .06$. These results suggest that judgements of environmental complexity were similar across both tasks, but that judgements of musical complexity were significantly better when participants were engaged in the visuospatial search task compared to when they were engaged in the language comprehension task. This pattern of results appears to reflect a syntactic interference effect. There was also a significant difference between judgements of environmental complexity when participants were completing the language task, and judgements of musical complexity when they were completing the visuospatial task, $t(36) = 4.91$, $p' < .001$, $d = .83$. There was no difference between judgements of musical complexity when performing the language task, and judgements of environmental complexity when performing the visuospatial search task, $t(36) = .91$, $p' = .74$, $d = .16$. Overall, this pattern of results shows that judgements of musical complexity were poorer when participants were engaged in the language comprehension task compared to when they were engaged in the visuospatial search task. The results suggest that interference occurred between musical complexity judgements and the language comprehension task that did not occur for environmental complexity judgements, suggesting a syntactic interference effect.

When WMC was included as a covariate in the above ANOVA measuring complexity judgements, the main effect of task became non-significant, $F(1, 35) = .77$, $p = .39$, the main effect of auditory stimulus became non-significant, $F(1, 35) = .08$, $p = .78$, and the interaction was still non-significant, $F(1, 35) = .84$, $p = .37$. There were also no interaction effects with WMC,[9] and no between-subjects effect of WMC, $F(1, 35) = 1.49$, $p = .23$. This pattern of results suggests that WMC can account for the

[9] There was no interaction between task and WMC, $F(1, 35) = .03$, $p = .39$, between auditory stimulus and WMC, $F(1, 35) = .84$, $p = 37$, and no three-way interaction, $F(1, 35) = .07$, $p = 79$.

differences between complexity judgements, and suggests that WMC is intricately linked to syntactic processing and the syntactic interference effect.

Discussion

The results of Experiment 2 were consistent with those of Experiment 1 in that no reliable differences were observed across the auditory conditions for either the language comprehension task or the visuospatial search task. However, Experiment 2 revealed that judgements of auditory complexity were dependent on the primary task that participants were completing. Participants' musical complexity judgements were significantly poorer when they concurrently completed a language comprehension task compared to when they concurrently completed a visuospatial search task.

General Discussion

The combined results of Experiments 1 and 2 showed no differences between four different auditory conditions on either a language comprehension task or a visuospatial search task. This outcome is surprising, and invites further consideration of the nature of interactions between music and language, and the conditions under which interference can be observed. In both experiments, complexity was manipulated in music and environmental sounds, and performance was measured on two primary tasks: a language comprehension task and a visuospatial search task. In Experiment 1, the secondary task involved judging whether the auditory stimulus was music or environmental sounds. In Experiment 2, the secondary task involved a more challenging judgement about whether the auditory signal was basic or complex. Performance on the secondary task was at ceiling in Experiment 1, suggesting that participants allocated some attention to processing the auditory stimulus, but that it was an extremely easy task resulting in very accurate performance. Experiment 2 showed that complexity judgements were affected by the concurrent primary task (language comprehension or visuospatial search), suggesting that interference was observed from the primary task to the secondary task. Furthermore, WMC was correlated with better performance on a number of language comprehension and visuospatial search measures, as would be expected. Interestingly, WMC was able to account for the syntactic interference observed in the secondary complexity judgements in Experiment 2. This finding suggests that working memory plays an important role in syntactic processing, and contributes to the syntactic interference effect.

Contrary to our findings, we predicted that complex music would result in reduced language comprehension as a result of syntactic interference, as observed in Fedorenko et al. (2009) and Patston and Tippett (2011). However, there were some crucial differences in these studies compared to the current experiments. Both previous studies involved out-of-key elements. Fedorenko et al. (2009) aligned an out-of-key note with an important word in a relative clause for both complex and basic sentences, which resulted in reduced comprehension accuracy for complex sentences. Patston and Tippett (2011) had participants complete as many language comprehension questions as possible in an eight-minute period, while listening to correct music, music with out-of-key notes throughout, or in silence. Participants also completed as many visuospatial searches as possible (the same stimuli as the current experiments) in an eight-minute period with the same auditory conditions. Patston and Tippett (2011) found that musicians showed the poorest performance on the language comprehension task when listening to music with out-of-key elements, and that correct music also resulted in fewer questions answered correctly than silence. This pattern is reflective of a syntactic interference effect; however, it was only observed in musicians, who could be argued to be particularly sensitive to musical stimuli. There were no differences between auditory stimulus conditions in the visuospatial search task for either musicians or nonmusicians. The study by Patston and Tippett (2011) showed syntactic interference when out-of-key elements were not aligned with the language. However, they only found syntactic interference in their musician sample, not in their non-musician sample. The experiments presented in this chapter included participants regardless of musical ability; however, preliminary analyses revealed no differences depending on musical training.

It appears that interference between music and language in a language comprehension task is observed when out-of-key notes are aligned with specific points in a sentence (Fedorenko et al., 2009), or judgements by musicians over a longer time period are measured (Patston & Tippett, 2011). However, a number of other published studies have measured performance on a language comprehension task, and found that language comprehension was not affected by auditory condition. Slevc et al. (2009) reported on a self-paced reading task that resulted in slower reading times for syntactically complex garden path sentences when they were combined with an out-of-key note. Participants were also asked comprehension questions at the conclusion of each trial, and no difference was observed for sentence comprehension depending on auditory condition. Therefore, in the study by Slevc et al. (2009), aligning an out-of-key note with a difficult point in the sentence did not result in lowered comprehension. A study aiming to replicate Slevc et al. (2009) using semantic garden path sentences also asked a comprehension question after each sentence (Perruchet & Poulin-Charronnat, 2013). These data are reported in a table but are not analysed, so it is assumed that there was no effect of auditory condition on language comprehension in this study either. It should be noted that in the study that did observe an effect on language comprehension, the auditory stimulus had to be sped up by 50% to find an effect, and there was no interference observed for self-paced reading times (Fedorenko et al., 2009).

The behavioural measurement of language comprehension therefore appears to be relatively robust to manipulations of syntactic structure in music. In an fMRI study by Kunert et al. (2015) basic and complex sentences were aligned with an out-of-key note or an increase in loudness on an important word in the sentence. Although they observed overlapping brain activation for music and language syntax processing, they did not observe the predicted behavioural difference in language comprehension depending on condition. It appears that the syntactic interference effect can only be observed with extremely sensitive measures—specifically, in ERP studies (Koelsch et al., 2005; Steinbeis & Koelsch, 2008), neuroimaging studies (Kunert et al., 2015), or using sensitive behavioural measures, such as self-paced reading (Slevc et al., 2009), or language recall (Fiveash et al., under review; Fiveash & Pammer, 2014). Language comprehension only appears to be affected by music syntax if the language stimuli are difficult to understand (Fedorenko et al., 2009), or the participants are musicians and the interference is prolonged (Patston & Tippett, 2011). From these results, it appears that the brain is able to quickly recover from any syntactic interference that occurs between music and language, and language comprehension tends to not be affected by manipulations of syntactic structure in music.

An interesting finding from Experiment 2 was that complexity judgements of auditory stimuli differed depending on the concurrent task the participant was engaged in. When concurrently engaged in the visuospatial search task, participants were significantly better at judging complexity in music compared to judging complexity in environmental sounds. However, when engaged in the language comprehension task, there was no difference between the musical and environmental complexity judgements. Furthermore, participants were significantly better at judging musical complexity when completing the visuospatial task compared to when they were completing the language task, whereas environmental complexity judgements did not differ depending on the concurrent task. These results suggest that the language comprehension task may have made it more difficult to perform the secondary task of judging complexity in music. Therefore, although interference was not observed in the primary, language comprehension task, it was observed in the secondary task of complexity judgements.

Conclusion

The experiments presented in this chapter showed that language comprehension was not affected by manipulations of complexity in either musical or environmental sounds. This result suggests that language comprehension is relatively robust to interference by background auditory stimuli. However, Experiment 2 revealed that musical complexity judgements were less accurate when participants were engaged in the language comprehension task compared to when they were engaged in the visuospatial search task, and that this difference did not occur for judgements of environmental complexity. Together, these results suggest that (a) language comprehension is robust against syntactic and complexity manipulations in accompanying auditory stimuli, and (b) syntactic interference can be observed in a secondary task.

Acknowledgments. I would like to thank Glenn Schellenberg for input on the design of Experiment 1, and for access to facilities and resources at University of Toronto, Mississauga for testing participants. Thank you to Shayan Alam for help testing participants for Experiment 1. Thank you also to Evelina Fedorenko and Lucy Patston for providing me with their stimuli.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Caplan, D. (2014). Commentary on "The role of domain-general cognitive control in language comprehension" by Fedorenko. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00629

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163-183.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172-179. doi:10.1016/j.tics.2010.01.004

Eerola, T. (2016). Expectancy-violation and information-theoretic models of melodic complexity. *Empirical Musicology Review, 11*(1), 1-17. doi:10.18061/emr.v11i1.4836

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology, 5*. doi:10.3389/fpsyg.2014.00335

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences, 108*(39), 16428-16433. doi:10.1073/pnas.1112937108

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences, 110*(41), 16616-16621. doi:10.1073/pnas.1315235110

Chapter 4: Complexity

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition, 37*(1), 1-9. doi:10.3758/MC.37.1.1

Fiebach, C., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during pasing: The processing of German WH-questions. *Journal of Memory and Language, 47*, 250-272.

Fiveash, A., McArthur, G., & Thompson, W. F. (under review). Music and language: Syntactic interference without syntactic violations.

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music, 42*(2), 190-209. doi:10.1177/0305735612463949

Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: Syntactic and semantic structures. *Frontiers in Psychology, 2*, 112. doi:10.3389/fpsyg.2011.00112

Hocking, J., Dzafic, I., Kazovsky, M., & Copland, D. A. (2013). NESSTI: Norms for environmental sound stimuli. *PLoS One, 8*(9), e73382. doi:10.1371/journal.pone.0073382

Koelsch, S., Gunter, T., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and music: An ERP study. *Journal of Cognitive Neuroscience, 17*(10), 1565-1577.

Koelsch, S., & Jentschke, S. (2008). Short-term effects of processing musical syntax: An ERP study. *Brain Research, 1212,* 55-62. doi:10.1016/j.brainres.2007.10.078

Chapter 4: Complexity

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS One, 10*(11), e0141069. doi:10.1371/journal.pone.0141069

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Loui, P., Grent-'t-Jong, T., Torpey, D., & Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Cognitive Brain Research, 25*(3), 678-687. doi:10.1016/j.cogbrainres.2005.08.019

Mathworks (R2016b). Matlab. Massachusetts, United States: Mathworks.

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

Patston, L. M., & Tippett, L. J. (2011). The effect of background music on cognitive performance in musicians and nonmusicians. *Music Perception, 29*(2), 173-183. doi:10.1525/mp.2011.29.2.173

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin & Review, 20*(2), 310-317. doi:10.3758/s13423-012-0344-5

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science, 22*(11), 1359-1366. doi: 10.1177/0956797611417632

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review, 16*(2), 374-381. doi:10.3758/16.2.374

Chapter 4: Complexity

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition, 116*(1), 71-86. doi:10.1016/j.cognition.2010.04.002

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18*(5), 1169-1178. doi:10.1093/cercor/bhm149

Tillmann, B., & Bigand, E. (2015). A commentary on "A commentary on: 'Neural overlap in processing music and speech'". *Frontiers in Human Neuroscience, 9*. doi:10.3389/fnhum.2015.00491

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - Fourth Edition*. San Antonio, TX: Pearson.

Chapter 5

Effects of Language Syntax on Music Syntax Processing

The research described in Chapter 4 revealed no reliable difference in language comprehension depending on manipulations of complexity or syntax in auditory stimuli. These findings suggest that language comprehension is relatively robust to concurrent auditory presentation. However, syntactic interference was observed in musical complexity judgements. Specifically, concurrent reading of sentences resulted in reduced performance on a secondary musical task. Experiment 3 in Chapter 2 also showed evidence that reading sentences resulted in poorer detection of out-of-key notes in a secondary music task compared to reading word-lists. This finding suggests that syntactic interference can be observed from language to music when music is the secondary task.

The experiments in Chapter 5 were primarily designed to determine whether language-to-music syntactic interference can also be observed when the music task is the primary task, in line with Question 3: can syntactic interference be observed from language to music? As mentioned in the Introduction, task prioritisation may play a crucial role in how syntactic processing resources are allocated to music and linguistic input, and very little research has addressed this issue. In Experiment 1, participants were asked to read sentences with (a) no error, (b) a semantic error, or (c) a syntactic error while performing a same-different judgement on two simple melodies. Participants then had to indicate whether or not there was a language error. Experiment 2 was a melody recall task, where participants listened to a melody played either alone or concurrently with a spoken sentence that contained (a) no error, (b) a semantic error, or (c) a syntactic error. They then had to indicate whether there was an error in the language stimuli. One of the outcomes from Experiments 1 and 2 was that semantic errors in sentences were detected with less accuracy than syntactic errors. To determine whether the lower detection of semantic errors was inherent to the stimuli or a result of the dual-task, Experiment 3 was designed to test how accurately participants detected syntactic and semantic errors when these errors were presented in a single-task situation. For Experiments 1 and 2, the music task was the primary task, and the language task was the secondary task.

This manuscript was co-authored by myself, Genevieve McArthur, Glenn Schellenberg, and Bill Thompson. I contributed approximately 85% of the total work, including experimental design, stimuli creation, experiment programming, data collection, statistical analysis, and the preparation of the first draft of the manuscript. Glenn Schellenberg and Bill Thompson contributed to the design and interpretation of Experiment 1. Genevieve McArthur and Bill Thompson contributed to the design of Experiment 2, the interpretation of results, and provided critical comments on the manuscript. Shayan Alam, a research student from the University of Toronto, Mississauga, tested approximately one third of the participants in Experiment 1.

This chapter was prepared as:

Fiveash, A., McArthur, G., Schellenberg, G., & Thompson, W. F. (under preparation). Effects of language syntax on music syntax processing.

Effects of Language Syntax on Music Syntax Processing

Anna Fiveash[1,3], Genevieve McArthur[2,3], Glenn Schellenberg[4], & William F. Thompson[1,3]

[1]Department of Psychology, Macquarie University

[2]Department of Cognitive Science, Macquarie University

[3]ARC Centre of Excellence in Cognition and its Disorders, Macquarie University

[4]Department of Psychology, University of Toronto, Mississauga

Chapter 5: Language to Music Interference

Music and language are two hierarchical systems that allow an infinite number of complex structures to be created from a finite number of smaller elements (e.g., notes, words). These complex structures (e.g., sentences, musical phrases) are built based on rules or "governing principles" of combination (Patel, 2008). It has been suggested that the brain processes incoming hierarchical information in a similar way for language and music, through a syntactic processing resource that is limited in capacity. The shared syntactic integration resource hypothesis (SSIRH) encapsulates this idea that *syntactic processing networks* are shared by music and language. It further postulates that *representation networks* store information specific to each domain (Patel, 2003, 2008). The SSIRH focuses primarily on overlap in *integrational* processes, and predicts that when music and language are taxing syntactic resources simultaneously, a cognitive cost should be evident. The syntactic equivalence hypothesis (SEH) also predicts that interference will occur when music and language syntax are concurrently processed (Koelsch, 2013). The SEH focuses on early overlap in initial structure building stages in addition to later integrational processes. Although both the SSIRH and the SEH predict that interference should occur when music and language syntax are processed simultaneously, they do not specify in which direction this interference might occur. The majority of studies investigating the SSIRH and SEH focus on interference from music to language processing. Though not explicitly stated in either theory, overlapping processing resources imply that interference should also be observed in the opposite direction: from language to music.

Previous research has shown processing costs when music and language syntax are processed concurrently. However, much of this research (especially behavioural research) has examined the effects of music on language processing, and for the majority of experiments, participants were told to ignore the accompanying music. For example, Slevc, Rosenberg, and Patel (2009) conducted a self-paced reading task, where garden path sentences were presented at the same time as individual chords. Participants read sentences that were presented one fragment at a time. As soon as they read and comprehended each fragment, they pressed a button in order to advance to the next fragment. The speed with which they progressed to the end of each sentence was recorded. There was no music task. Slevc et al. (2009) found that participants were significantly slower at reading ambiguous sections of garden path sentences when they were paired with an out-of-key chord compared to when those fragments were paired with an in-key chord. Importantly, reading time was not affected when the sentence contained a semantic anomaly, suggesting a syntax-specific effect (though see Perruchet & Poulin-Charronnat, 2013). Fedorenko, Patel, Casasanto, Winawer, and Gibson (2009) also conducted a self-paced reading task where each word in a sentence (either a complex or simple sentence) was sung on a different note of a melody. The task was to answer comprehension questions about the sentences, and again, there was no music task. Fedorenko et al. (2009) found that comprehension accuracies were significantly worse for complex sentences when a structurally important word was sung on an out-of-key note, compared to the simple sentences. This effect did not occur when the same note was sung at a different loudness level (auditory anomaly), suggesting that the effect is syntax-specific, and not related to attention-grabbing events.

Hoch, Poulin-Charronnat, and Tillmann (2011) investigated syntactic interference by simultaneously presenting chords that were either expected (tonic chord) or unexpected (subdominant chord) in the tonal context at the same time as either syntactically expected or unexpected final words in a lexical decision task. Chord sequences were aligned with each syllable in a sentence, and participants had to decide whether a final target word was a word or a non-word as quickly and as accurately as possible. Hoch et al. (2011) found the typical *syntactic expectancy effect;* that is, syntactically expected words were classified as words more quickly than syntactically unexpected words. They also found the typical *tonic facilitation effect*—syntactically expected words were classified as words more quickly when they were presented with an expected chord compared to an unexpected chord. Crucial for this experiment, Hoch et al. (2011) also found an interaction between syntactic expectancy and tonal function. Specifically, the syntactic expectancy effect was evident when the target word was accompanied by the tonic chord, but decreased when it was accompanied by the subdominant chord. Furthermore, the tonic facilitation effect was not observed for syntactically unexpected words.

Hoch et al. (2011) conducted the same experiment with semantically unexpected words instead of syntactically unexpected words. This experiment provided evidence for the *semantic expectancy effect* (semantically expected words were processed faster than semantically unexpected words) and the tonic facilitation effect as expected. However, there was no interaction between semantic expectancy and tonic facilitation. This pattern of results suggests that music syntax and language syntax interfere across domains, whereas music syntax and language semantics do not interact. As in Slevc et al. (2009) and Fedorenko et al. (2009), music accompanied the language task, but participants were not asked to make judgements on the music. In all of the studies mentioned above, a music syntax manipulation presented at the same time as a language syntax manipulation resulted in poorer or slower performance on a language task. However, as music processing was not measured, it cannot be ascertained whether language processing interfered with music processing.

The same is true for two further studies (Fiveash, McArthur, & Thompson, under review; Fiveash & Pammer, 2014). Fiveash and Pammer (2014) measured recall of written complex sentences (with syntax) compared to recall of written word-lists (no syntax) whilst participants were listening to background music that had (a) no violation, (b) a syntactic violation (out-of-key chord), or (c) a timbre violation (a chord played with a different timbre). Recall of sentences was significantly worse when paired with music with a syntactic violation compared to music containing a timbre change or no violation. This interference was not observed for recall of word-lists, suggesting that the out-of-key chord specifically interfered with syntactic processing in sentences, and not in other forms of language processing.

Fiveash et al. (under review) aimed to investigate sentence and word-list recall without out-of-key elements, to limit both sensory violations and salient attention grabbing events (see Tillmann & Bigand, 2015). Fiveash et al. (under review) observed interference in recall of written sentences when participants were listening to background melodies compared to environmental sounds (Experiment 1). This pattern did not occur in written word-lists, suggesting a syntax-specific effect. In a subsequent experiment (Experiment 3), participants directed their attention to either a melody or a sequence of environmental sounds while reading and then recalling word-lists and sentences. Participants were asked to detect targets that occurred in 20% of trials (an out-of-key note in melodies, and a gong sound in environmental sounds). Fiveash et al. (under review) reported more errors in sentence recall when participants were attending to the melodies compared to environmental sounds. Furthermore, detection of out-of-key notes was significantly reduced when participants were concurrently reading sentences compared to word-lists. Gong detection was not affected depending on concurrent language presentation, suggesting that the syntax in the sentences interfered with the music task of detecting out-of-key notes. This study provides preliminary support for an effect of language processing on music judgements when music is the secondary task.

Koelsch, Gunter, Wittfoth, and Sammler (2005) measured event-related potentials (ERPs) elicited with the simultaneous presentation of written sentences and chord sequences. The sentences had either a syntactic or a semantic violation, and the chord sequences ended on either an expected or unexpected chord. Participants were told to ignore the music, and to respond on 10% of trials if they detected a syntactic or semantic error in the sentences. Koelsch et al. (2005) found that the neural response to a syntactic violation in language, the left anterior negativity (LAN), was significantly reduced when the language syntax violation was paired with a music syntax violation. However, they found no reduction of the N400 neural response to the semantic violation, suggesting this effect is specific to syntax. They also found that the brain response to the music syntax violation, the early right anterior negativity (ERAN), was not modulated by a language syntax violation. The authors suggested that processing the music violation drew resources away from processing the language violation, and that this finding reflects shared, limited-capacity syntactic processing resources.

All of the studies described above—each showing an effect of music syntax on language syntax processing—used either no music task or a secondary music task (e.g., target detection). Without a primary music task, it is difficult to determine if the reverse effect of interference from language syntax to music syntax processing was present. However, research using dual-task paradigms provide evidence that violations of language syntax affect the processing of music syntax. For example, Steinbeis and Koelsch (2008) presented participants with chord sequences ending on either the tonic or an unexpected chord, and sentences that contained no error, a syntactic error, or a semantic error. Participants attended to the music to detect random timbre deviants, and to the language to answer memory questions spread throughout the experiment. A syntactic violation in music produces both the ERAN ERP at an early stage of processing, and the N5 ERP at a later stage of processing. The N5 is suggested to reflect processes of musical meaning (Koelsch, 2005; Steinbeis & Koelsch, 2008). When participants attended to both music and language, the ERAN response to a music syntax violation was significantly reduced when paired with a language syntax violation, but not a semantic violation. This finding suggests that interference can occur from language to music. Semantic violations and music syntax violations interacted at the later N5 stage, supporting the separation of the ERAN and N5 components for syntactic and semantic processing. Corroborating results from Koelsch et al. (2005), Steinbeis and Koelsch (2008) also found that the LAN was reduced when paired with an unexpected compared to an expected chord. These results show syntactic interference from music to language, and from language to music, but only when participants are asked to pay attention to both domains.

To the author's knowledge, apart from the secondary target detection task in Experiment 3 in Fiveash et al. (under review), only two behavioural studies have shown effects of language syntax processing on music syntax processing. Kunert, Willems, and Hagoort (2016) presented participants with written sentences and chord sequences that were presented simultaneously, fragment by fragment. Chord sequences started in one key, contained a *pivot* chord that led into a second key, and were resolved with an authentic cadence completing either the first key or the second key. Sentences were either ambiguous garden path sentences that required a revision, or non-ambiguous sentences that did not require a revision. Participants were asked to judge how "complete" each chord sequence sounded (harmonic closure), and to answer a language comprehension question. The results showed that compared to non-ambiguous sentences, garden path sentences reduced harmonic closure ratings for music. However, this finding only occurred when the chord sequence was resolved in the first key, not in the second, suggesting an interaction with long distance integration. This finding is taken as evidence for an effect of language syntax on music syntax processing. Kunert et al. (2016) used an arithmetic task as a control, and found that increased difficulty in the arithmetic task did not alter harmonic closure ratings for either first or second key endings, again suggesting a syntax-specific effect. The finding that harmonic closure ratings decreased for first key endings when paired with garden path sentences was replicated in a follow-up experiment. The same pattern did not occur for semantic garden path sentences. Harmonic closure ratings in these experiments showed large ceiling effects, and only occurred with first key endings. However, the results are suggestive of an effect of language syntax processing on music syntax processing.

In a recent study by Van de Cavey, Severens, and Hartsuiker (2017), participants were presented with control sentences, garden path sentences (resolvable), and sentences with a syntactic violation (unresolvable) at the same time as pitch sequences (these were not melodies as they were not created to be tonal). Sentence segments were presented one at a time with a simultaneous pitch. Pitch sequences were created to include *cluster shifts*—shifts between groups of pitches arranged based on the circle of fifths (e.g., a C-F-G cluster compared to a B-E-A cluster). Shifts between different clusters were suggested to create phrase boundaries. After listening to the pitches and reading the words, participants were presented with two tones (e.g., C, G), and asked whether those tones were present in the same order in the previous pitch sequence. The tones presented in the probe tone task were manipulated to be (a) *between-probes*, where the test tones had been presented over a cluster shift in the pitch sequence; (b) *within-probes*, where the test tones had occurred within a cluster; or (c) *foil probes*, where the test tones had occurred in the pitch sequence, but not in the order presented in the probe tone task.

Van de Cavey et al. (2017) based their design on previous research that found poorer recognition of two tones when those tones were presented over a phrase boundary. Their rationale was that a difficult structural integration in language (e.g., a garden path sentence) would reduce the resources available to process the phrase boundary, eliminating the between-probe effect, and resulting in better recognition of tones. Van de Cavey et al. (2017) predicted that the same pattern should not occur with syntactic errors because syntactic errors cannot be resolved, and therefore should not draw on the same resources allocated to phrase boundary processing. Van de Cavey et al. (2017) found some evidence for this hypothesis. They found that participants performed better on the between-probe task when the tones were presented with a garden path sentence compared to a sentence with a syntactic violation. They suggested that this finding is evidence that language integration processes have an effect on the processing of music syntax, and argued that the overlap between music and language is at the level of integration, rather than specific to syntax. Further research needs to be conducted with this paradigm to assess its validity, as the effects were quite small, and the pitch sequences were not musical as such. This finding also contradicts previous research that has found interference with syntactic violations in language (e.g., Hoch et al., 2011; Koelsch et al., 2005; Steinbeis & Koelsch, 2008).

There is contradictory evidence regarding the types of stimuli that cause syntactic interference, suggesting that there is considerable confusion in the field about the precise processing resources that may be shared between music and language. Patel (2008) suggested that the shared processes are related to syntactic integration costs of activating representations in long-term memory. Koelsch (2013) suggested that in addition to integrational processes, music and language also interact at an earlier, structure building stage of processing. Based on evidence that interference was found for syntactic (Slevc et al., 2009) and semantic (Perruchet & Poulin-Charronnat, 2013) garden path sentences but not for semantic anomalies, Slevc and Okada (2014) proposed that the shared resources are related to cognitive control. However, it is unclear whether cognitive control processes are activated when there is no conflict to resolve, as in naturalistic music and language processing. The maintenance, activation, and online integration of incoming music and language sequences appear to utilise many of the processes suggested by Baddeley and Hitch (1974) in the multi-store model of working memory, especially in relation to the central executive and the episodic buffer (Baddeley, 2000). Furthermore, syntactic processing in both music and language appears to engage working memory resources (Burunat, Alluri, Toiviainen, Numminen, & Brattico, 2014; Fiebach, Schlesewsky, & Friederici, 2002; King & Just, 1991; Kljajević, 2010; Vos, Gunter, Kolk, & Mulder, 2001). To measure the contribution of working memory to syntactic processing in the current experiments, working memory capacity (WMC) was measured for each participant.

The current experiments were designed with three aims in mind. The first aim was to investigate whether the syntactic processing of language interferes with the syntactic processing of music under dual-task conditions. The second aim was to investigate whether language processing interferes with music processing when the music stimulus does not contain out-of-key elements that introduce added sensory violations (Koelsch, 2013; Tillmann & Bigand, 2015). The third aim was to investigate the effect of WMC on concurrent music and language processing. In Experiment 1, written sentences were presented with (a) no error, (b) a semantic error, or (c) a syntactic error at the same time as the first melody in a melodic same-different task. Experiment 2 was designed to examine whether recall for melodies was influenced by spoken sentences with (a) no error, (b) a semantic error, or (c) a syntactic error. For both experiments, it was predicted that sentences with a syntactic error should interfere with the music task more than sentences with a semantic error or no error. These predictions were made based on theories that music and language share limited capacity online syntactic structure building resources (Patel, 2008; Koelsch, 2013). A grammatical error in a sentence places a burden on the processing of that sentence, resulting in syntactic reanalysis (Friederici, 2002). As the sentence requires more syntactic resources, there are fewer resources available for online structure building in music, and the resulting encoded information is likely to be reduced. The reduction of encoded information should then result in poorer performance on the same-different and melody recall tasks. Semantic errors do not place this structural burden on resources, and are therefore not expected to interfere to the same level as the syntactic errors. In Experiment 3, detection of semantic and syntactic errors in sentences was measured in a single-task paradigm.

Working memory capacity and musical training were also measured. It was predicted that WMC and musical training would be positively correlated with performance on the musical memory tasks (same-different and melody recall), given that higher WMC should facilitate recall for music stimuli, and people with musical training should have greater familiarity with musical tasks. Further, it was predicted that WMC would account for potential syntactic interference effects because of the importance of working memory to syntactic processing. Based on previous research it was also predicted that syntactic interference would be more evident in participants with musical training (Patston & Tippett, 2011). In line with the SSIRH and SEH, if music and language share syntactic processing resources, then an effect of language syntax on music syntax processing should be observed when the two types of stimuli are concurrently processed.

Experiment 1

Method

Participants. Sixty-four participants (51 females, 13 males, M_{age} = 18.5 years, *SD* = 2.7) from the University of Mississauga, Canada, participated for course credit. Of these, 53 listed English as their first language. Of the 11 who listed English as their second language, all learnt English before the age of eight. Second languages varied widely, with 34 participants indicating that they spoke another language fluently, and 30 indicating that they only spoke English. Eleven participants indicated that they had five or more years of private music training, and in total participants had an average of 2.2 years of private music training (*SD* = 3.62). Ten participants indicated that they were currently musically active, and 62 reported listening to music every day (M_{minutes} = 228, *SD* = 176).

Design. Experiment 1 was a 2 (melodies: same, different) by 3 (sentences: no error, semantic error, syntactic error) within-subjects design. There were 45 *same* trials, and 45 *different* trials that were paired with 30 normal (no error) sentences, 30 sentences with a semantic error, and 30 sentences with a syntactic error. Each type of sentence was paired with 15 same and 15 different trials. There were 90 trials in total. Presentation was randomised for each participant, and it was ensured that the same melody was not presented in succession. Stimuli were presented through headphones in a sound proof room with Matlab (version 2016b; Mathworks) and Psychtoolbox (Brainard, 1997).

Melodies. Melodies were created from the top line notes of the chord sequences from Slevc et al. (2009), which were all strongly tonal and in the key of C. Melodies contained 8-11 notes (*M* = 9.36; 8 notes = 6, 9 notes = 22, 10 notes = 12, 11 notes = 5). To create 45 distinct melodies, 21 of the melodies were produced backwards, while maintaining the last two notes in their final position. One melody was not used as it contained four of the same notes in a row. Stimuli never had more than three notes repeated in succession. All stimuli were transposed to the key of A, and created using the musical instrument digital interface (MIDI) instrument Steinway grand piano in GarageBand. Initially, to create the *different* trials, the order of two adjacent notes was reversed, resulting in a contour change. However, pilot testing showed that this manipulation was too difficult, and produced very low detection scores on the same-different task. Therefore, in the experiment proper, *different* trials were created by reversing the order of the last four notes (excluding the final two notes). For example, the melody C$^{\#}$-A-B-**C$^{\#}$-B-A-A-**G$^{\#}$-A became C$^{\#}$-A-B-**A-A-B-C$^{\#}$-**G$^{\#}$-A. All melodies were 5 seconds long. As such, tempo was altered depending on the amount of notes: 8 notes = 96 beats per minute (bpm), 9 notes = 108bpm, 10 notes = 120bpm, 11 notes = 132bpm.

Language stimuli. Language stimuli were taken from Osterhout and Nicol (1999). There were 30 normal sentences (no error), 30 sentences with a semantic error, and 30 sentences with a syntactic error. The errors were always located in the middle of the sentence. An example of a normal (no error) sentence is: *Alison used a hammer to break the small lock*. An example of a semantic error is: *Alison used a hammer to* kiss *the small lock*, and an example of a syntactic error is: *Alison used a hammer to* breaking *the small lock*. Sentences averaged 11.7 words (range: 8-16 words) and were randomly allocated into the three conditions. No sentences were repeated. An independent-samples *t*-test ensured there was no difference in word-count between the three conditions.

Procedure. All participants signed the information and consent form and completed a short questionnaire of demographic and musical training questions. Participants were informed about the study and completed practice trials consisting of two trials with only the music task, and two trials with the music and language task combined. Participants heard and were informed that melodies had different tempi, and were given examples of syntactic and semantic errors in language. It was reinforced that both types of examples were considered a language error. The experimenter ensured the participant understood the task, and the experiment proper began. On each trial, a fixation-cross first appeared on the screen for 1 second. The first melody then played through headphones at the same time as the written sentence was presented on a white screen. The first melody and sentence were concurrently presented for 5 seconds. A grey screen then appeared, and the second melody played after 2 seconds. Participants indicated via they keyboard whether the melodies were the same (press z) or different (press m). To ensure participants focused on the melodies as the primary task, they received feedback (correct or incorrect) on the music task only, as pilot testing showed that feedback increased performance. After indicating whether the melodies were the same or different, participants indicated if there had been an error in the sentence: yes (press z) or no (press m). To ensure participants read the whole sentence, they also had to indicate what the last word in the sentence was. Participants were given a forced choice of either the last word, or a random word that was the last word from another sentence. The position (left or right) of the correct word was randomised. The language tasks were therefore the secondary tasks. This procedure was repeated 90 times, with a break every 20 trials. See Figure 1.

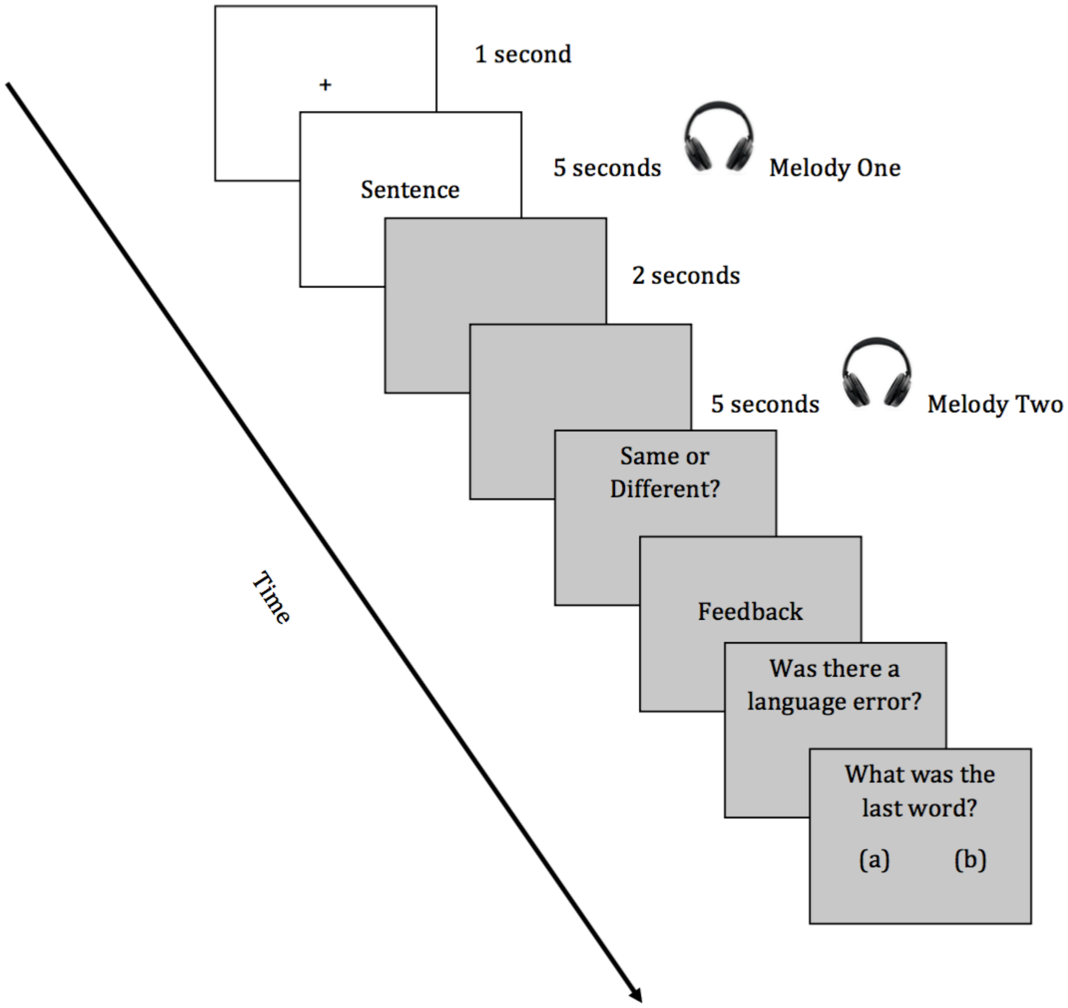

Figure 1. Procedure for Experiment 1. Boxes reflect the computer screen, and the timing on the right indicates how long each screen appeared for. Headphones represent the presentation of melodies.

Following the experiment proper, participants completed the letter-number sequencing sub-test of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 2008). The letter-number sequencing sub-test is a measure of WMC, and was included to investigate the role of WMC in syntactic processing. The test was presented via the computer. Participants were instructed that they would view a sequence of numbers and letters, each presented for 1 second on the screen, followed by the # sign. They were told to rearrange the numbers and letters in their head so that the numbers occurred first in ascending order, followed by the letters in alphabetical order (e.g., 9-C-3 should be re-ordered as 3-9-C). Participants had five practice trials, consisting of two- and three-element combinations of numbers and letters. The experimenter ensured the participant understood the concept and was answering correctly in the practice trials. Participants were told to manipulate the sequence in their head, and to only write down the final answer. The test consisted of three trials each of 2, 3, 4, 5, 6, and 7 letter-number sequences, resulting in 21 trials, and a WMC score out of 21. The whole testing session took approximately 50 minutes.

Scoring and analysis. D prime (d') scores for the primary same-different music task were calculated based on signal detection theory (Stanislaw & Todorov, 1999). D prime is a measure of sensitivity to the signal, in terms of hits and false alarms. The z score for false alarm rate (indicating the melodies were different when they were the same) was subtracted from the z score for hit rate (correctly indicating the melodies were different), resulting in the d' score. Extreme hit or false alarm values (e.g., 1 or 0) were corrected for: scores of 1 were replaced by (n – 0.5) and scores of 0 were replaced by 0.5/n, where n = the number of signal or noise trials. See Stanislaw and Todorov (1999) for more detail.

Accuracy and reaction time (RT) scores were calculated for the secondary language tasks. For sentence error detection, participants were correct if they correctly identified a sentence error, or correctly identified no sentence error. D prime scores were also calculated as above by using the normal (no error) condition as the false alarm rate, and the semantic and syntactic conditions as hit rates. Accuracy and RT scores for the last word sentence judgements were also calculated.

The analysis consisted of repeated measures analyses of variance (ANOVAs) for the music same-different judgements, the last word judgements, the sentence judgements, and the RTs for each of these judgements. Working memory capacity (as measured by performance on the letter-number test of the WAIS) was added as a covariate into the above ANOVAs to investigate effects of WMC on performance. Musical training (measured by years of private lessons) was also added into the above ANOVAs as a covariate to investigate the effect of musical training on melodic same-different judgements and sentence error detection. For brevity, the influence of musical training on performance is reported in footnotes; however, full statistical analyses of musical training are reported in Appendix B. ANOVAs are reported first without covariates, and then with covariates, as recommended by Simmons, Nelson, and Simonsohn (2011). Between-subjects effects were investigated with bivariate correlations, and any significant main effects or interactions were investigated with paired-samples *t*-tests. All Cohen's *d* values reported are based on repeated measures calculations, which take into account paired-samples correlations between variables.

Results

Melodic same-different judgements. A repeated measures ANOVA was conducted on the d' scores for the melodic same-different task across the three language conditions (no error, semantic error, syntactic error). There was no main effect of condition; same-different judgements did not differ depending on whether the first melody was paired with a normal (no error) sentence ($M = .82$, $SD = .73$), a semantic error sentence ($M = .90$, $SD = .70$), or a syntactic error sentence ($M = .94$, $SD = .69$), $F(2, 126) = .31$, $p = .74$. However, when WMC was added as a covariate, the main effect of language condition was significant, $F(2, 124) = 3.55$, $p = .032$, $\eta^2 = .05$. There was also a significant interaction between language condition and WMC, $F(2, 124) = 3.30$, $p = .04$, $\eta^2 = .05$, and a significant between-subjects effect of WMC, $F(1, 62) = 6.48$, $p = .01$, $\eta^2 = .10$. These effects suggest that WMC significantly affected performance on the same-different task, and that the effect of WMC differed depending on condition.[1]

Reaction times for the musical same-different task did not differ depending on condition, $F(1.64, 103.12) = .60$, $p = .52$ (the Greenhouse-Geisser correction was reported as the assumption of sphericity was violated, $\chi^2(2) = 15.55$, $p < .001$). When WMC was added as a covariate into the RT ANOVA, the main effect of language condition remained non-significant, $F(1.64, 101.67) = .95$, $p = .39$, and there was no interaction between condition and WMC, $F(1.64, 101.67) = .78$, $p = .46$. However, there was a significant between-subjects effect of WMC, $F(1, 62) = 4.25$, $p = .04$, suggesting that WMC affected RT on the melodic same-different task.[2]

To investigate the above interactions and between-subjects effect of WMC, bivariate correlations were performed between WMC and both the same-different d' scores and same-different RTs in each condition. Working memory capacity was positively correlated with performance on the same-different task when the melody was paired with normal (no error) sentences ($r = .33$, $p = .007$) and semantic error sentences ($r = .34$, $p = .006$), but not with syntactic error sentences ($r = .044$, $p = .73$; see Figure 2). The same pattern was observed for the RT data. WMC was negatively correlated (participants with higher WMC answered more quickly) with RT for the normal (no error) sentences ($r = -.26$, $p = .04$) and semantic error sentences ($r = -.27$, $p = .03$), but not for the syntactic error sentences ($r = -.19$, $p = .14$). These correlations may suggest that for the normal (no error) and semantic error conditions, participants had to read the

[1] An analysis presented in Appendix B that included musical training as a covariate in the ANOVA examining d' scores revealed a significant between-subjects effect of musical training. Bivariate correlations showed that participants with more musical training performed better on the musical same-different task, as would be expected. There was still no main effect of language condition, and no interaction between language condition and musical training.
[2] When musical training was added as a covariate into the same-different RT ANOVA, there was still no main effect of language condition, no interaction between language condition and musical training, and no between-subjects effect of musical training.

whole sentence to detect an error, thereby placing a greater burden on working memory. Because the syntactic error was immediately obvious, and in the middle of the sentence, it is possible that participants stopped reading the sentence as soon as they detected the syntactic error, and therefore performance was not as dependent on WMC in the syntactic error condition.

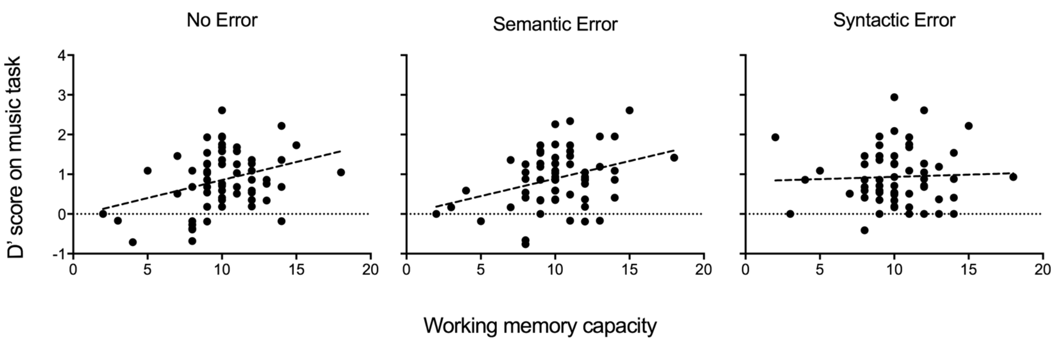

Figure 2. Correlations between working memory capacity (WMC) and d' scores on the melodic same-different task depending on language condition. WMC was measured by the letter-number sequencing sub-test of the Wechsler Adult Intelligence Scale. Possible scores range from 0-21.

The above interpretation that participants paid less attention to the sentence as soon as they detected a syntactic error was investigated by analysing accuracy of last word judgements. Reduced recognition of the last word in the syntactic error condition would suggest that participants were less likely to read or pay attention to the end of the sentence. To analyse last word judgements depending on language condition, a repeated measures ANOVA was performed. This analysis revealed a main effect of language condition on last-word judgements, $F(1.48, 92.97) = 13.64$, $p < .001$, $\eta^2 = .18$. The assumption of sphericity was violated, $\chi^2(2) = 27.21$, $p < .001$, so Greenhouse-Geisser corrected values have been reported. Pairwise comparisons (Holm-Bonferroni corrected for three comparisons) revealed that participants were significantly worse at correctly indicating the last word in the sentence in the syntactic error condition ($M_{accuracy}$ = .91, *SD* = .09) compared to both the no error ($M_{accuracy}$ = .95, *SD* = .05), *t*(63) = 3.93, *p'* < .001, *d* = .53, and semantic error ($M_{accuracy}$ = .95, *SD* = .06) conditions, *t*(63) = 4.0, *p'* < .001, *d* = .52. There was no difference between the no error and the semantic error conditions, *t*(63) = .00, *p'* = 1.0. When WMC was included as a covariate, the main effect of language condition was non-significant, *F*(1.48, 91.50) = 1.36, *p* = .26, there was no interaction between language condition and WMC, *F*(1.48, 91.50) = .07, *p* = .94, and there was a significant between-subjects effect of WMC, *F*(1, 62) = 5.76, *p* = .019, η^2 = .09. Bivariate correlations revealed that WMC was positively correlated with last word judgements in the no error condition (*r* = .31, *p* = .01) and the semantic error condition (*r* = .25, *p* = .04), but not the syntactic error condition, (*r* = .22, *p* = .09). These results are compatible with the correlations between WMC and melodic same-different judgements in the no error and semantic error conditions, and the lower accuracy for last word judgements in the syntactic condition, and further suggest that the syntactic error condition resulted in task switching to the music task.

Participants were also significantly slower at last word judgements in the syntactic error condition ($M_{seconds}$ = 1.76, *SD* = .52) compared to the no error condition ($M_{seconds}$ = 1.59, *SD* = .42), *t*(63) = 4.31, *p'* < .001, *d* = .56. However, they were significantly slower at last word judgements in the semantic error ($M_{seconds}$ = 1.69, *SD* = .45) compared to the no error condition as well, *t*(63) = 2.5, *p'* = .03, *d* = .31. There was no difference in RT for last word judgements in the syntactic error compared to the semantic error condition, *t*(63) = 1.76, *p'* = .08 (Holm-Bonferroni corrected *p* values were reported for three comparisons). The last word judgement accuracy scores suggested that participants were more likely to stop reading the sentence to focus on the melody in the syntactic error condition, thereby freeing up resources for music processing. Reaction time results showed that both the syntactic and semantic error conditions resulted in slower reaction times for last word judgements.

Error detection in language. A repeated measures ANOVA was conducted to examine accuracy on the language error detection task across the three language conditions (no error, semantic error, syntactic error). The assumption of sphericity was violated, $\chi^2(2) = 10.43$, $p = .005$, so a Greenhouse-Geisser correction was applied. The ANOVA revealed a significant main effect of language condition, $F(1.73, 109.1) = 7.47$, $p = .002$, $\eta^2 = .11$. Planned pairwise comparisons with Holm-Bonferroni adjusted *p* values for three comparisons showed that participants were significantly worse at detecting semantic errors ($M_{\text{accuracy}} = .72$, $SD = .20$) compared to correctly identifying no error ($M_{\text{accuracy}} = .80$, $SD = .14$), $t(63) = 2.48$, $p' = .03$, $d = .32$, or syntactic errors ($M_{\text{accuracy}} = .82$, $SD = .17$), $t(63) = 4.38$, $p' < .001$, $d = .52$. There was no difference between the no error and the syntactic error conditions, $t(63) = .78$, $p' = .44$. A paired-samples *t*-test comparing the d' values for the semantic and syntactic error conditions also showed that participants were more sensitive to syntactic errors ($M = 2.03$, $SD = .80$) compared to semantic errors ($M = 1.62$, $SD = .75$), $t(63) = 4.77$, $p < .001$, $d = .60$. The finding that participants were significantly worse at detecting the semantic errors compared to syntactic errors and no errors suggests that semantic errors are less noticeable than syntactic errors in a dual-task situation.

To investigate whether WMC affected sentence error detection, the same repeated measures ANOVA was performed to examine accuracy scores for the language error detection task with the added covariate of WMC.[3] The assumption of

[3] When musical training was added as a covariate into the sentence error detection ANOVA, there was still a main effect of language condition, no interaction between musical training and language condition, and no between-subjects effect of musical training, suggesting that musical training did not influence error detection in language.

sphericity was violated, $\chi^2(2) = 10.37$, $p = .006$, so the Greenhouse-Geisser correction has been applied. When WMC was included as a covariate, there was still a main effect of language condition, $F(1.73, 107.24) = 3.50$, $p = .033$, $\eta^2 = .05$, and no interaction between language condition and WMC, $F(1.73, 107.24) = 2.96$, $p = .06$. However, there was a significant between-subjects effect of WMC, $F(1, 62) = 5.49$, $p = .02$, $\eta^2 = .08$.

To investigate the between-subjects effect of WMC and the marginal interaction effect, bivariate correlations were conducted for WMC and sentence error detection accuracy across the three sentence conditions (see Figure 3). The results indicated an opposite pattern to the correlations between WMC and melodic same-different judgements. WMC was significantly correlated with sentence error detection accuracy for the syntactic error condition ($r = .34$, $p = .005$), but not for the semantic error ($r = .22$, $p = .08$) or the no error ($r = -.03$, $p = .79$) conditions. Combined with the WMC correlations in the music task, these correlations could be interpreted to suggest that processing resources were the most engaged for the no error and semantic error conditions. The semantic error and no error conditions required a thematic analysis of the whole sentence. This thematic analysis may have resulted in correlations with WMC on the music task, as people with higher WMC were better able to manage the dual-task situation. Participants indicated whether there was a language error after hearing a second melody. Consequently, they had to remember their sentence judgement (error or no error) for at least five seconds, plus the length of time it took for them to respond to the music question. It is possible that participants noticed syntactic errors quickly, and switched their attention to the music task. The correlations with WMC and judgements of syntactic errors may therefore have occurred because participants with higher WMC were better able to remember their initial (quick) judgement.

Figure 3. Correlations between working memory capacity and accuracy on the sentence judgement task, where participants were asked if the sentence contained an error. For the no error condition, accuracy was based on participants indicating there was no error.

A repeated measures ANOVA performed on the RTs for language error detection revealed a main effect of language condition, $F(1.76, 110.55) = 7.49$, $p = .001$, $\eta^2 = .11$ (the Greenhouse-Geisser correction was applied as the assumption of sphericity was violated, $\chi^2(2) = 9.33$, $p = .009$). Pairwise comparisons with Holm-Bonferroni adjusted p values revealed that participants were significantly quicker at syntactic error judgements ($M_{\text{seconds}} = 1.69$, $SD = .76$) compared to judging there was no error ($M_{\text{seconds}} = 1.87$, $SD = .79$), $t(63) = 4.89$, $p' < .001$, $d = .59$. There were no differences in RT for syntactic or semantic error judgements ($M_{\text{seconds}} = 1.79$, $SD = .77$), $t(63) = 1.89$, $p' = .126$, and no difference between the no error and the semantic error conditions, $t(63) = 1.67$, $p' = .13$. To summarise, participants reported syntactic errors significantly faster than no errors, but there was no difference in RT between reporting syntactic and semantic errors. When WMC was added as a covariate into the above ANOVA, there was no main effect of language condition, $F(1.78, 109.88) = 2.40$, $p = .10$, no interaction between language condition and WMC, $F(1.78, 109.88) = 1.72$, $p =$

.19, and no between-subjects effect of WMC, $F(1, 62) = .61$, $p = .44$, suggesting that WMC accounted for the variance in RT across conditions.[4]

Discussion

Experiment 1 revealed no difference in melodic same-different judgements depending on whether participants were reading sentences with no error, a syntactic error, or a semantic error. Experiment 1 therefore revealed circumstances under which syntactic interference was not observed. Specifically, it appeared that interference was not observed when the participant had time to recover (e.g., the second melody was played without concurrent language stimuli), or when full sentences with errors were presented, as written sentences may have resulted in task switching and effects of reading that eliminated the syntactic interference effect. Participants with musical training performed better on the same-different task, as predicted, but this effect did not differ depending on language condition. Further, it was observed that WMC affected the processing of concurrent musical sequences and written sentences, and it appears that the salience of the language error may have affected how processing resources were allocated to each task in the dual-task.

The current pattern of results suggests that participants may have been task switching (switching attention between tasks) to process melodies and sentences simultaneously. Performance on the melodic same-different task was correlated with WMC for the semantic error and no error conditions, but not the syntactic error condition. These correlations suggest that once participants encountered a syntactic error, they switched their attention to the melody, therefore limiting the impact of differences in WMC. On the other hand, the semantic error and no error conditions

---

[4] When musical training was added as a covariate into the sentence error detection RT ANOVA, there was still a main effect of language condition, no interaction between musical training and language condition, and no between-subjects effect of musical training. These results suggest that musical training does not influence RT for language error detection.

required a full analysis of the sentence in working memory to determine whether there was an error. This interpretation is supported with two lines of evidence. First, last word judgements were significantly worse in the syntactic error condition compared to both the no error and semantic error conditions, suggesting participants may not have read the whole sentence as carefully in the syntactic error condition. Second, sentence error judgements were only correlated with WMC for the syntactic error condition. This pattern of results may suggest that participants made a quick and less considered judgement about syntactic errors. To recall this judgement after processing a second melody and making a decision on the same-different task may have required more working memory resources for the syntactic error condition, leading to this correlation.

The current results allow an important insight into task prioritisation in a dual-task situation. These data suggest that as soon as participants were able to make a decision about one stream of information, they could switch attention to the other stream to perform as accurately as possible on both tasks. The finding that there was no difference on the musical same-different task (the primary task) depending on sentence condition suggests that participants were particularly resilient to the current dual-task paradigm. Future research should limit task switching, either by presenting words in a self-paced reading task, or by presenting spoken sentences so participants are forced to process both streams simultaneously.

Two aspects of the current design help to elucidate the conditions under which syntactic interference is not observed between music and language. The first aspect relates to the sentence stimuli. The normal sentence (no error) condition may have led to an exhaustive search, as participants had to keep analysing the sentence to determine if there was an error. This exhaustive search may have required more processing resources than sentences with errors. Further, because errors were located in the middle of the sentences, participants could effectively switch their attention to the melody once they had detected the error. To limit this possibility, participants were asked to read the whole sentence beginning to end, and to indicate the final word in each sentence. However, it still appears that task switching occurred. Considering syntactic errors were detected with higher accuracy than semantic errors, it is possible that this difference eliminated any effects of the language errors on music processing. The second aspect relates to stimulus presentation. Only the first melody was presented with language. The second melody was played alone. This design may have resulted in the release of any potential strain on syntactic processing resources, and a recovery from possible interference. Future research investigating syntactic interference should therefore either use stimuli with no language errors, or place errors carefully within the sentence to limit task switching and exhaustive search. Furthermore, future research should also ensure that interference paradigms are taxing both streams simultaneously at all times, to be sensitive enough to observe interference.

Experiment 2 was designed to address these suggestions in three ways. First, the primary task was changed from a melodic same-different with concurrent written stimuli to a melody recall task with concurrent spoken stimuli. The melody recall task mirrors previous research investigating sentence recall when listening to music, and should be a more sensitive measure of syntactic processing (Fiveash et al., under review; Fiveash & Pammer, 2014). Second, instead of reading sentences while listening to music, in Experiment 2, participants listened to spoken sentences and music simultaneously. Spoken sentence stimuli remove effects of reading, and therefore limit task switching. Further, the simultaneous presentation of spoken sentences and melodies is hypothesised to increase the burden on syntactic processing resources between music and language. Third, to eliminate the issue with presenting an error in the middle of a sentence, the stimuli in Experiment 2 contained syntactic and semantic errors at the end of each sentence. Sentences with no errors should result in distributed attention across the whole sentence, and sentences with errors on the last word should limit task switching. The enforced, simultaneous processing of music and speech is predicted to result in syntactic interference effects that can be observed in the recall of melodies.

Experiment 2

Experiment 2 was designed to investigate whether syntactic errors in spoken sentences interfere with the recall of melodies. Previous research has paired out-of-key notes or chords with syntactic and semantic errors in language to show interference from music to language (e.g., Hoch et al., 2011; Koelsch et al., 2005). To investigate whether syntactic violations in language affect music processing in the same way that syntactic violations in music affect language processing, participants listened to six-note melodies presented alone, or concurrently with (a) a sentence with no error (baseline), (b) a sentence with a semantic error, or (c) a sentence with a syntactic error. To ensure no effect of task switching (as observed in Experiment 1), errors always occurred on the last word of the sentence.

The melody recall task was based on research by Williamson, Baddeley, and Hitch (2010). In a number of experiments, Williamson et al. (2010) tested recall for melodies of different lengths (3-8 notes), different distances between notes (proximal versus distal), and different numbers of distinct notes (numbers not reported, but a pilot study mentioned in the paper suggested an optimal number of three). The results of these experiments suggested that melody recall decreased when (a) the number of notes increased, (b) notes were close in pitch, or (c) there were more than three distinct pitches within a melody, particularly for non-musicians. Based on these findings, recall for seven-note melodies was initially tested in a pilot study presented in Appendix A. This pilot study revealed that seven-note melodies were very difficult for participants to recall when simultaneously listening to sentences, and six-note melodies were subsequently used to minimise the chance of both ceiling and floor effects. These melodies contained three distinct notes that were seven semitones (a perfect fifth) apart. It was predicted that syntactic errors in sentences would reduce recall for melodies compared to semantic errors and no errors, as the syntactic error in the sentence stimuli should interfere with structure building processes and disrupt the memory trace for the melody representation (Koelsch, 2013; Patel, 2008).

Method

Participants. Fifty-two undergraduate psychology students from Macquarie University participated for course credit (M_{age} = 21.5 years, *SD* = 6.9; 44 females, 8 males). All participants spoke English as their first language, and all but two were born in an English-speaking country. The large majority had parents who spoke English as their first language (mother: 85% reported English as first language, father: 81%). Forty-eight were right-handed, and all reported listening to music, with an average listening time of 135 minutes per day (*SD* = 80 minutes). None reported any suspected or diagnosed hearing or language issues. Forty participants reported some formal musical training, with an average of 5.7 years (*SD* = 6) of private training, and 9 years (*SD* = 10.4) of combined private and informal experience. Fourteen participants reported being currently musically active, four considered themselves a musician, 38 did not consider themselves a musician, and 10 “somewhat” considered themselves a musician. Two participants reported having perfect pitch.

Design. Experiment 2 investigated the effects of concurrent speech stimuli on melody recall. Melody recall was investigated under four conditions: (a) melody only,

(b) concurrent normal sentence (no error), (c) concurrent sentence with a semantic error, and (d) concurrent sentence with a syntactic error. Eighty-eight unique melodies were created and randomly paired with one of the conditions. Though pairings were constant across participants, two different randomisations were created and counterbalanced across participants to limit effects of specific pairings. The experiment was programmed and presented with Matlab (version R2016b, Mathworks) and Psychtoolbox (Brainard, 1997).

Melodic stimuli. Based on results of a pilot study (Appendix A), six-note melodies were created. The 96 unique ways that three notes can be combined without immediate repetition were created using Matlab. Melodies that did not contain all three notes were excluded, resulting in 90 possible melodies. Two of these melodies were presented in the practice trials, leaving 88 unique melodies consisting of the notes C_4 (the tonic note), G_4 (the fifth), and D_5 (the second, in the next octave). An equal proportion of melodies started on each note. MIDI scores were created for each melody in MuseScore (version 2.0.2), imported into GarageBand, and played through the MIDI instrument Steinway grand piano. Melodies were 140bpm to equal the length of the speech stimuli. Each melody was therefore slightly over 2.5 seconds long.

Speech stimuli. Spoken sentences were from Sun, Lu, Ho, Johnson, and Thompson (2015). Sentences were spoken by a female, Australian, native-English speaker with natural and clear prosody. The sentences were all in the form: *Nancy is teaching one girl*, or *Nancy is teaching two girls*. A syntactic error involved a number disagreement, for example: *Nancy is teaching one girls*, or *Nancy is teaching two girl*. A semantic error included a word that did not make sense in the context, for example: *Nancy is teaching two bags*, or *Nancy is teaching one bag*. An equal amount of singular and plural sentences were randomly selected from Sun et al. (2015), and all errors occurred on the last word. No sentences were repeated, and semantic errors were specifically chosen to be as anomalous as possible considering the poorer detection of semantic errors in Experiment 1. All sentence recordings were approximately 2-3 seconds long, and mostly 2.5 seconds. To ensure consistent presentation with melodic stimuli, sentence recordings were altered so that each sentence was exactly 2.5 seconds long. Each sentence was checked after altering to ensure no artificial sounds were introduced.

Stimulus pairings. Melodies and sentences were normalised to an equal loudness level with Audacity. Melodies were randomly allocated to a condition (melody only, no error, semantic error, or syntactic error) and within each condition were randomly allocated to a sentence. Two different randomisations were created to limit effects of specific pairings. Melody-sentence pairs were created in Audacity and exported as mp3 files. The alignment of the sentences and melodies meant that the semantic or syntactic error was presented at approximately the same time as either the last two notes, or the last note, depending on whether the last word was plural or singular.

Procedure. Participants completed an information and consent form and a musical education questionnaire and then received pitch training. Pitch training consisted of listening and pitch judgement exercises. First, participants heard a C chord, followed by the three notes (C, G, D) they would be discriminating. These notes were described as “low”, “medium”, and “high”. The three notes were played 10 times in a row to familiarise participants with the distances between each note. Participants then had 10 trials where they heard a note and indicated whether it was low, medium, or high. For these trials, participants first heard the C chord, and then the target note. Feedback on this task indicated whether they were correct or incorrect. If incorrect, the correct answer was provided. If participants were not confident in this task they could repeat the process. The procedure was explained to participants, and they practiced the melody recall task two times using the 2, 5, and 8 keys on the right keypad of the keyboard to indicate the low, medium, or high notes. Participants were told that some of the melodies were paired with speech, and that the speech may or may not contain an error. It was made clear that the error could either be a grammatical error or a meaning error, and participants were asked to indicate whether they heard an error. There were three melody recall practices when the melody was presented at the same time as the sentences, and participants heard examples of each sentence type: no error, semantic error, and syntactic error. After the practice trials, the experimenter ensured the participant understood the task, and the experiment proper began. For each trial, a C chord was played, followed by the three notes in the melodies (C, G, D) played at 120bpm. The stimuli then played for its duration (approximately 2.5 seconds). Once the stimuli finished, the word “recall” appeared on the screen, and participants recalled the order of the notes using the keypad. They were then asked “Was there a language error?” and responded with yes (press 3) or no (press 1). See Figure 4. Participants responded with the hand that felt most comfortable to them (usually the right). There were 88 trials, with a break every 22 trials. The experiment took approximately half an hour. After the experiment proper, participants completed the letter-number sequencing sub-test of the WAIS, as in Experiment 1.

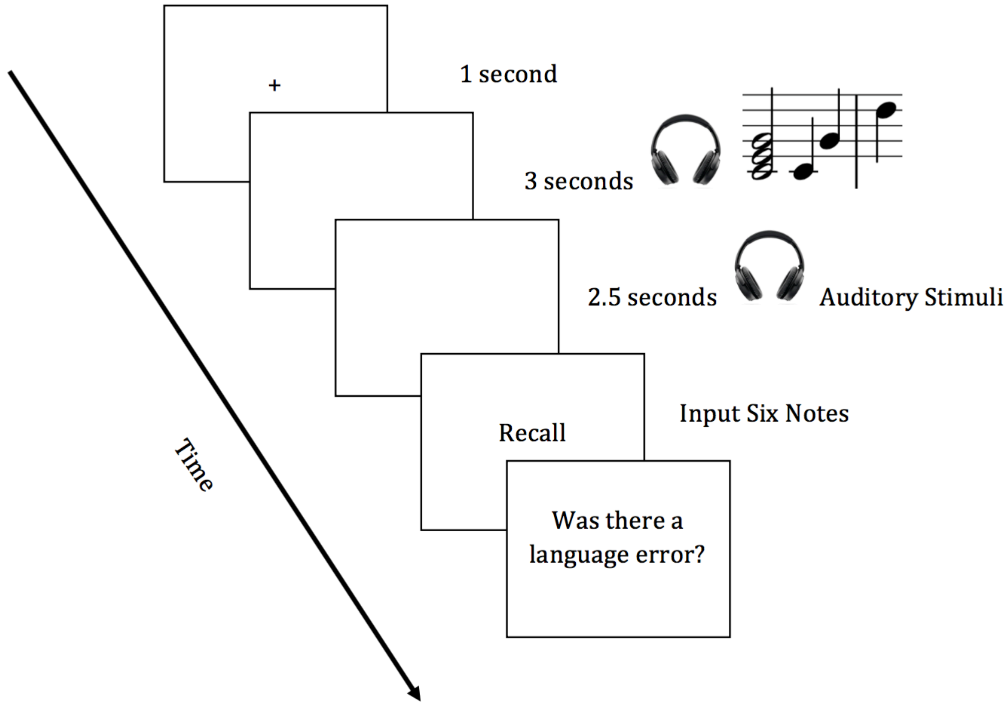

Figure 4. Procedure and timing for Experiment 2. Boxes represent the computer screen, and headphones indicate auditory stimuli.

Scoring and analysis. Melody recall was first scored as absolute values across each note position (1-6) based on whether the participant indicated the correct note in the correct position. Average accuracy across all positions was then calculated to create a total recall score across the whole melody. Accuracies were calculated for each condition: melody only, concurrent normal sentence (no error), concurrent sentence with a semantic error, and concurrent sentence with a syntactic error. Analysis using the total accuracy scores from these data is first presented, reflecting recall for *absolute pitch*. However, an exploratory analysis (not reported here) showed that starting note (C, G, or D) affected melody recall in the first position, and starting note was not evenly distributed among conditions in each randomisation. To eliminate this confound, and to focus on an important element of melody recall—the ability to recall *relative pitch*—the melodies were then scored for relative pitch accuracy. Absolute pitch judgement scores are reported to provide a complete picture of the data.

To calculate relative pitch accuracy, correct identification of the interval between each note was scored. The three possible notes were the “low” C, the “medium” G, and the “high” D. For example, if the melody went up from G to D, this interval would reflect a step of +1. If the melody went down from D to C, this interval would reflect a step of -2. Melody recall was therefore analysed in terms of whether participants correctly indicated the relative pitch, regardless of absolute pitch. If the notes were C to G (+1), and the participant responded G to D (+1), they would score correctly, as their judgement of relative pitch was correct. The accuracy of relative pitch judgements for each position (from 2-6) was scored, resulting in a score for each position. Data were also averaged across all positions for a total melody recall score for each condition. These data were used for overall analyses, and position analyses for each condition. Accuracy on the error detection task for the sentences was also measured, as were d’ scores using the normal sentences as the false alarm rate.

Repeated measures ANOVAs were conducted to examine melody recall across the whole sequence for both the absolute pitch scoring and the relative pitch scoring. Relative pitch scores are insensitive to starting note and sensitive to recall for relative pitch, and are therefore most reflective of true differences between conditions. The relative pitch scores were therefore used in all analyses after the preliminary reporting of absolute pitch scores. Melody recall as a function of position and condition was analysed in a repeated measures ANOVA, and further analyses investigated evidence for a recency effect in immediate serial recall. Working memory capacity and musical training (years of private music lessons) were included as covariates throughout the analysis. For brevity, musical training results are presented in footnotes, and full analyses are reported in Appendix B. All results of ANOVAs are reported first without covariates, and then with covariates, as suggested by Simmons et al. (2011). All

Cohen's *d* values reported are based on repeated measures calculations, which take into account paired-samples correlations between variables.

Results

Absolute pitch analysis. A repeated measures ANOVA examined the absolute, total melody recall across the four conditions: melody only, no error, semantic error, and syntactic error. This analysis revealed a significant main effect of condition $F(3, 153) = 11.31$, $p < .001$, $\eta^2 = .18$. Paired-samples *t*-tests with Holm-Bonferroni adjusted *p* values for six comparisons revealed that the melody only condition ($M = .57$, $SD = .16$) resulted in significantly better recall than all other conditions: no error ($M = .51$, $SD = .16$), $t(51) = 4.36$, $p' < .001$, $d = .61$; semantic error ($M = .53$, $SD = .14$), $t(51) = 3.26$, $p' = .008$, $d = .46$; and syntactic error ($M = .50$, $SD = .14$), $t(51) = 5.56$, $p' < .001$, $d = .79$. The semantic error condition also resulted in significantly better melody recall than the syntactic error condition, $t(51) = 2.64$, $p' = .03$, $d = .37$. There was no difference between the no error and semantic error condition, $t(51) = 1.22$, $p' = .46$, or the no error and syntactic error condition, $t(51) = .98$, $p' = .46$.

To investigate the role of WMC in melody recall, and whether this differed depending on condition, the same repeated measures ANOVA was conducted with the added covariate of WMC. When WMC was taken into account, there was still a main effect of condition, $F(3, 150) = 3.0$, $p = .03$, $\eta^2 = .06$, and no interaction between condition and WMC, $F(3, 150) = .90$, $p = .44$. There was a between-subjects effect of WMC, $F(1, 50) = 8.1$, $p = .006$, $\eta^2 = .14$, suggesting that WMC influenced melody recall results, but the effect of WMC did not differ depending on condition.[5]

---

[5] When musical training was added as a covariate into the ANOVA with absolute pitch scores, there was still a main effect of condition, no interaction between condition and musical training, and a between-subjects effect of musical training. Bivariate correlations revealed that musical training was positively correlated with melody recall across all conditions.

Correlations between WMC and performance on the melody recall task showed that WMC was positively correlated with performance in each condition: melody only ($r = .28$, $p = .048$), no error ($r = .34$, $p = .01$), semantic error ($r = .34$, $p = .01$), and syntactic error ($r = .44$, $p = .001$), suggesting that participants with higher WMC also performed better on the melody recall task.

Relative pitch analysis. To investigate whether the same pattern occurred in the relative pitch judgements (eliminating effect of starting note) compared to the absolute pitch judgements, the same analysis was conducted with the relative pitch scoring for total melody recall. Similar to the absolute judgements, a repeated measures ANOVA on the relative pitch judgements across the four conditions showed a main effect of condition, $F(3, 153) = 6.4$, $p < .001$, $\eta^2 = .11$. Paired-samples *t*-tests (with Holm-Bonferroni adjusted *p* values reported for six comparisons) again showed that the melody only condition ($M = .49$, $SD = .17$) resulted in significantly better recall than all other conditions: no error ($M = .44$, $SD = .17$), $t(51) = 3.76$, $p' < .001$, $d = .52$, semantic error ($M = .45$, $SD = 16$), $t(51) = 3.14$, $p' = .01$, $d = .44$, and syntactic error ($M = .44$, $SD = .16$), $t(51) = 4.1$, $p' < .001$, $d = .57$. However, in contrast to the absolute values, there was no difference between the semantic and syntactic error conditions in melody recall, $t(51) = .75$, $p' = 1.0$. This analysis suggests that when starting note and relative pitch judgements are taken into account, there is no difference between the semantic error and syntactic error conditions. In line with the absolute pitch judgements, there were still no differences between the no error and semantic error conditions, $t(51) = .41$, $p' = 1.0$, or the no error and syntactic error conditions, $t(51) = .26$, $p' = 1.0$. See Figure 5.

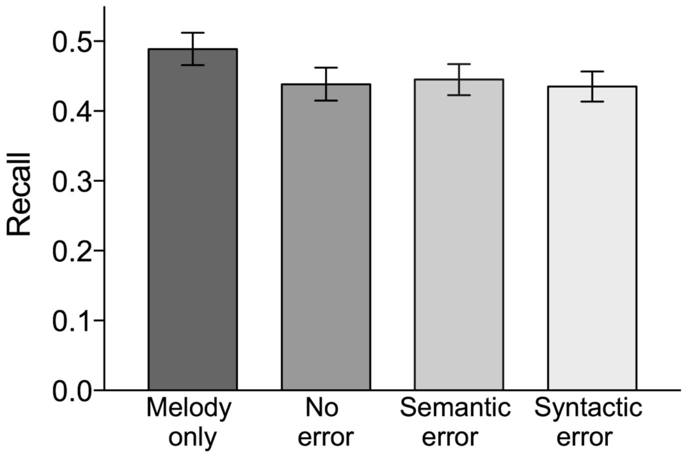

Figure 5. Melody recall scored with relative pitch judgements. Recall is measured as average proportion correct across the five note positions for each condition. Error bars indicate one standard error either side of the mean.

The same repeated measures ANOVA with WMC as a covariate showed that when WMC capacity was taken into account, the main effect of condition was non-significant, $F(3, 150) = 1.0$, $p = .39$. This finding suggests that WMC accounts for a sizeable amount of variance between conditions, and eliminates the difference between the melody only condition compared to the concurrent melody and sentence conditions. There was no interaction between condition and WMC, $F(3, 150) = 21$, $p = .89$, suggesting that WMC did not alter performance depending on condition. However, there was a between-subjects effect of WMC, $F(1, 50) = 11.03$, $p = .002$, $\eta^2 = .18$, suggesting that WMC influenced recall. Correlations between WMC and the relative pitch recall scores in each condition showed that WMC was significantly positively correlated with all conditions: melody only ($r = .36$, $p = .008$), no error ($r = .41$, $p = .003$), semantic error ($r = .38$, $p = .006$), and syntactic error ($r = .43$, $p = .001$), as in the absolute pitch scoring. These correlations show that participants with higher WMC performed better on the melody recall task, as would be expected on a recall task.[6]

Melody recall as a function of position. To investigate whether recall in each position (2-6) differed depending on condition, a 4 (condition: melody only, no error, semantic error, syntactic error) by 5 (position: 2-6) repeated measures ANOVA was conducted using the relative pitch scores. There was a main effect of condition, $F(3, 153) = 6.4$, $p < .001$, $\eta^2 = .11$, a main effect of position, $F(4, 204) = 44.08$, $p < .001$, $\eta^2 = .46$, and an interaction between condition and position, $F(12, 612) = 4.5$, $p < .001$, $\eta^2 = .08$. See Figure 6. These results suggest that condition affected melody recall differently across the different note positions. When WMC was added as a covariate, the main effect of condition was non-significant, $F(3, 150) = 1.0$, $p = .39$, the main effect of position was non-significant, $F(4, 200) = 1.7$, $p = .15$, and the interaction between condition and position was still significant, $F(12, 600) = 2.39$, $p = .005$, $\eta^2 = .05$. These results suggest that WMC can account for the variance between conditions and across positions, but that the interaction between condition and position remains. Further, there was no condition by WMC interaction, $F(3, 150) = .21$, $p = .89$, and no position by WMC interaction, $F(4, 200) = 1.53$, $p = .20$. However, there was a significant three-way interaction between condition, position, and WMC, $F(12, 600) = 1.85$, $p = .038$, $\eta^2 = .036$. There was also a significant between-subjects effect of WMC, $F(1, 50) = 11.03$, $p = .002$, $\eta^2 = .18$. These results suggest an important influence of

[6] Including musical training as a covariate into the melody recall ANOVA shows that musical training can also account for much of the variance between conditions (the main effect of condition was reduced to $F(3, 150) = 2.66$, $p = .050$). There was no condition by musical training interaction, but there was a between-subjects effect of musical training, again showing that participants with more musical training performed better on the melody recall task.

WMC in the recall of a melodic sequence, and that this influence differs depending on concurrent language condition.[7]

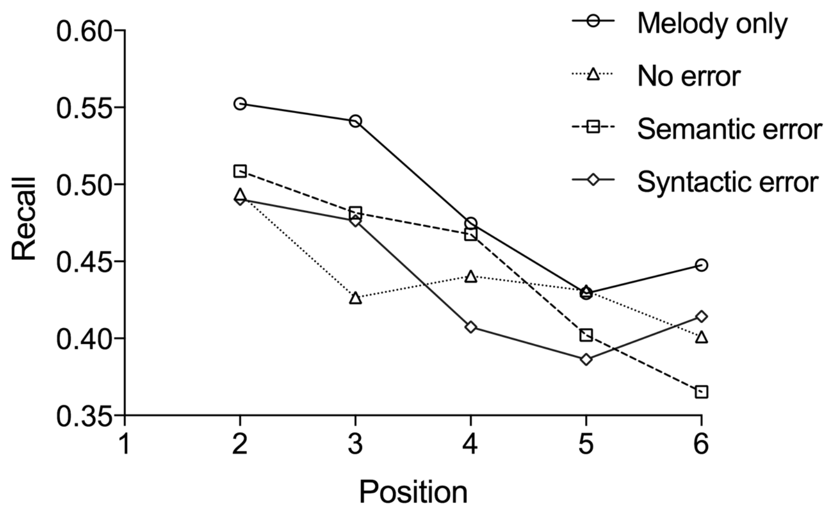

Figure 6. Melody recall for relative pitch judgements in each position, in each condition. Position 1 has no data, as relative pitch judgements were based on whether participants correctly indicated the direction of each note change, not absolute values.

A visual appraisal of the position data in Figure 6 suggests that from the penultimate to the final position, sentence recall increased for the melody only and syntactic error conditions, but decreased for the semantic error and no error conditions. The increase for the final note in the melody only and syntactic error conditions appears to reflect the recency effect, where performance tends to be better at the end of a to-be recalled sequence than the middle (Crowder & Greene, 2000; Roberts, 1986; Rundus & Atkinson, 1970). Serial recall usually shows a "bow" shape, where performance is best at the beginning, worst in the middle, and then recovers toward the end of a sequence. This pattern has been well documented in melody recall studies (Roberts, 1986) as well

[7] When musical training was added as a covariate into the condition by position ANOVA, the main effect of condition was non-significant (p = .050), the main effect of position was still significant, and the interaction between position and condition was still significant. There were no interaction effects. These results suggest musical training did not affect melody recall across positions, or the interaction between condition and position.

as other studies of immediate serial recall (Glanzer & Cunitz, 1966; Gupta, 2005; Jahnke, 1963, 1965). Considering the effect of the first note was removed with the relative pitch scoring, only the recency effect will be examined. In addition to the visual appraisal, recall in the final two note positions is important to analyse as the language errors occurred in these positions.

To investigate whether condition affected the recency effect, a 4 (condition: melody only, no error, semantic error, syntactic error) by 2 (position: 5, 6) repeated measures ANOVA was conducted to examine whether there was a difference between the fifth and sixth positions as a function of condition. This analysis revealed a main effect of condition, $F(3, 153) = 3.79$, $p = .01$, $\eta^2 = .07$, no main effect of position, $F(1, 51) = .45$, $p = .51$, but an interaction between condition and position, $F(3, 153) = 3.87$, $p = .01$, $\eta^2 = .07$, suggesting that the recency effect differed depending on condition. When WMC was included as a covariate, the main effect of condition was non-significant, $F(3, 150) = 1.7$, $p = .17$, the main effect of position was significant, $F(1, 50) = 4.94$, $p = .03$, $\eta^2 = .09$, and the interaction between condition and position was not significant, $F(3, 150) = .75$, $p = .53$. There was also a significant position by WMC interaction, $F(1, 50) = 6.17$, $p = .016$, $\eta^2 = .11$, but no three-way interaction between condition, position, and WMC, $F(3, 150) = 1.39$, $p = .25$. In addition, there was a significant between-subjects effect of WMC, $F(1, 50) = 7.22$, $p = .01$, $\eta^2 = .13$. These results suggest that WMC can account for the differences between conditions in relation to the recency effect and that WMC is inherently tied to performance on the melody recall task.[8]

[8] When musical training was added as a covariate into the ANOVA investigating recall in the last two positions, there was no main effect of condition ($p = .050$), no main effect of position, but an interaction between condition and position. There were no interactions.

To further investigate the condition by position interaction observed when WMC was not controlled for, difference scores were calculated by subtracting the 5th position recall scores from the 6th position recall scores, resulting in positive scores when recall increased in the final position, and negative scores when recall decreased in the final position (see Figure 7).

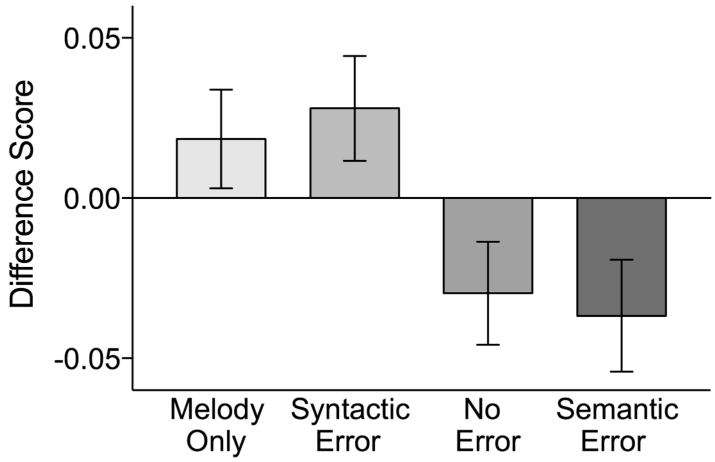

Figure 7. Melody recall difference scores in each condition for the final position (6th) minus the 5th position. Positive scores reflect increased recall from the 5th to 6th position, and negative scores reflect decreased recall from the 5th to 6th position. Error bars indicate one standard error either side of the mean.

Paired-samples *t*-tests on the difference data (with Holm-Bonferroni adjusted *p* values for four comparisons) revealed that the melody only difference score (M = .02, SD = .11) was significantly different from the semantic error (M = -.04, SD = .13), t(51) = 2.7, p' = .04, d = .38 difference score, and marginally different to the no error (M = -.03, SD = .12), t(51) = 2.0, p' = .05, d = .28 difference score. These comparisons reflect the finding that the melody only score increased between the 5th and 6th position, whereas the semantic error and no error scores decreased. This finding is clearer when examining the increased recall in the syntactic error condition (M = .03, SD = .12), which was significantly different to the decreased recall for the semantic error, t(51) =

2.43, $p' = .048$, $d = .35$, and no error, $t(51) = 2.48$, $p' = .048$, $d = .34$ conditions. These data suggest a difference between processes involved in final position recall for melody only and syntactic error conditions compared to no error and semantic error conditions.

Sentence judgements. Participants were also asked to indicate if they detected an error in the sentences. A repeated measures ANOVA on error detection accuracy for the three sentence conditions (no error, semantic error, syntactic error) revealed a significant effect of condition, $F(1.18, 60.15) = 27.77$, $p < .001$, $\eta^2 = .35$. The assumption of sphericity was violated, $\chi^2(2) = 59.5$, $p < .001$, so the Greenhouse-Geisser correction was applied. Paired-samples *t*-tests on these data (Holm-Bonferroni adjusted *p* values reported for three comparisons) showed that participants were significantly worse at detecting errors in the semantic error condition ($M = .72$, $SD = .26$) compared to the syntactic error condition ($M = .92$, $SD = .08$), $t(51) = 5.35$, $p' < .001$, $d = 1.33$. Participants were also worse at detecting semantic errors than correctly identifying no error ($M = .94$, $SD = .06$), $t(51) = 5.47$, $p' < .001$, $d = .83$. There was no difference between correctly identifying no error, and identifying a syntactic error, $t(51) = 1.6$, $p' = .11$. See Figure 8. Poorer recognition of semantic errors compared to syntactic errors was also observed with d' scores using the normal sentences as the false alarm rate. Participants were significantly less sensitive to semantic errors ($M = 2.30$, $SD = .85$) compared to syntactic errors, ($M = 3.06$, $SD = .62$), $t(51) = 5.73$, $p < .001$, $d = .81$. When WMC was added as a covariate into the sentence error detection ANOVA, the main effect of condition was no longer significant, $F(1.18, 58.97) = 2.04$, $p = .14$, suggesting that WMC can account for the lower identification of semantic errors. There was no condition by WMC interaction, $F(1.18, 58.97) = .01$, $p = 1.0$, and no between-subjects effect of WMC, $F(1, 50) = .58$, $p = .45$.[9]

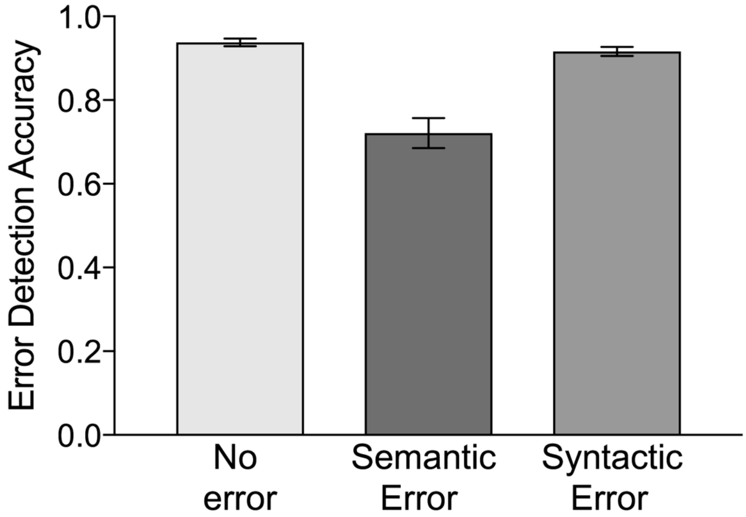

Figure 8. Accuracy in correctly identifying (a) no error, (b) a semantic error, and (c) a syntactic error in the sentence stimuli. Error bars indicate one standard error either side of the mean.

Discussion

The results of Experiment 2 indicated that recall for full melodies decreased significantly when paired with sentences. This decrease did not differ depending on whether the sentences contained a syntactic error, a semantic error, or no error, but was strongly related to WMC. An analysis investigating relative pitch recall across each position in the melodies showed a difference depending on condition. The melody only and syntactic error conditions showed signs of a recency effect in melody recall (an increase in recall from the penultimate to final position), whereas the no error and semantic error conditions resulted in decreased recall from the penultimate to final

[9] When musical training was added into the sentence error detection ANOVA, the main effect of condition was still significant, and there was no interaction between condition and musical training. However, there was a between-subjects effect of musical training, and bivariate correlations revealed that musical training was positively correlated with semantic error detection but not syntactic error detection or correct identification of no error.

position. These findings suggest that different cognitive processes were occurring for the semantic error and no error conditions compared to the syntactic error and melody only conditions. Further, musical training was positively correlated with performance on the melody recall task, suggesting that participants with more musical training were better at recalling melodies, as predicted.

Syntactic errors in the current stimuli introduced local errors, for example: *Nancy is teaching one girls*. To determine if there was a syntactic error, participants only needed to compare the last two words. As soon as they compared these words, and decided there was an error, they could give full attention to maintaining the melodic memory trace. On the other hand, to determine whether there was a semantic error or no error, participants needed to compare the final word with the representation built up online to decide whether it was semantically correct. For example, *Nancy is teaching two bags* requires the reader to reactivate the sentence memory trace to determine whether bags are able to be taught. The poorer performance in the final position for the semantic and no error conditions may therefore be related to the reactivation and semantic analysis of the sentence. This reactivation and semantic reanalysis may not occur with the type of syntactic agreement error used in the current experiment. This hypothesis is supported by previous eye-tracking data which showed that syntactic errors (occurring in the middle of the sentence) resulted in very quick and mandatory eye-movement reactions, whereas pragmatic (semantic) anomalies resulted in more regressive eye-movements towards the beginning of the sentence (Braze, Shankweiler, Ni, & Palumbo, 2002). Braze et al. (2002) also suggested that pragmatic anomalies showed prolonged differences in eye-movements compared to control sentences, and were processed after a delay, whereas syntactic information was processed quickly, and eye-movements recovered rapidly. Differences between semantic and syntactic processing in relation to the current results will be discussed further in the General Discussion.

A major finding of Experiment 2 was the contribution of WMC to the observed effects. WMC was measured because working memory is suspected to play an important role in syntactic processing. Overall, the results suggested that melody recall improved as WMC increased, as would be expected for a memory task. The results further suggested that WMC could account for the differences in recall between a single-task (melody recall) and a dual-task (melody recall and sentence error detection). WMC could also account for the difference in recall across each note position, and for the difference between conditions in the recency effect. There was also a complex interaction between WMC, condition, and position, suggesting that WMC is intricately tied to syntactic interference between music and language, and is likely to be a large part of the shared processing resources between music and language. The results also revealed that participants were worse at detecting semantic errors than syntactic errors in sentences. However, this difference was also accounted for by WMC, suggesting that the lower detection of semantic errors was related to the working memory resources available to each participant. It is therefore possible that the processing load in the dual-task situation accounted for reduced detection of semantic errors compared to syntactic errors and no error. Experiment 3 was designed to investigate this question.

Experiment 3

Based on the results of Experiments 1 and 2 it appears that the detection of semantic errors was worse than the detection of syntactic errors when these errors were paired with music. This observation is concerning, considering many previous studies have compared the influence of semantic and syntactic errors in the investigation of interactions between music and language syntax processing. It is therefore possible that syntactic errors are inherently more noticeable than semantic errors, and are more likely to show interference with music processing. In Experiment 3, sentence stimuli from Experiment 1 (written sentences) and Experiment 2 (spoken sentences) were presented to participants without concurrent music, to investigate error detection for semantic and syntactic errors in a single-task paradigm. If semantic errors are detected more poorly than syntactic errors, this finding would suggest that they are not appropriate comparison stimuli. However, if semantic and syntactic errors are detected at a similar level, this finding would suggest that semantic error detection is specifically hindered by concurrent music in a dual-task situation. For the written sentences in Experiment 1, semantic error detection was still poorer than syntactic error detection when WMC was controlled for. However, for the spoken sentences in Experiment 2, there was no difference between semantic and syntactic error detection when WMC was taken into account. It was therefore predicted that the reduced detection of semantic errors in Experiment 1 was related to the stimuli, whereas the reduced detection of semantic errors in Experiment 2 was related to the dual-task paradigm.

Method

Participants. Eighteen undergraduate psychology students from Macquarie University participated in the study for course credit. Two participants scored three *SD* below the mean on error detection in at least one of the conditions, so were excluded as outliers. This resulted in 16 participants (M_{age} = 20 years, *SD* = 1.5; 13 females, 3 males). Fifteen participants spoke English as their first language. One participant spoke Chinese as their first language. However, this participant was born in an English-speaking country, had their education in English, and did not perform differently than other participants, so was included in the analysis. Fifteen participants were born in an English-speaking country, and the majority had parents who spoke English as their first language (mother: 75% reported English as first language, father: 62.5%). All reported speaking English at home 60% or more of the time (11 reported 95% or more). None reported any suspected or diagnosed language disorders, and only one reported a slight (<5%) hearing decrement in one ear.

Design and stimuli. The experiment was a 2 (sentence type: written, spoken) by 3 (error type: no error, semantic error, syntactic error) repeated measures design. The written sentences from Experiment 1 and the spoken sentences from Experiment 2 were tested in this study. Written sentences consisted of the 30 sentences with no error, 30 sentences with a semantic error, and 30 sentences with a syntactic error from Experiment 1. All written sentences were devised by Osterhout and Nicol (1999). Errors occurred in approximately the middle of each sentence. Spoken sentences were 22 sentences with no error, 22 sentences with a semantic error, and 22 sentences with a syntactic error from Experiment 2. These were designed by Sun et al. (2015). Twenty-two more normal sentences from the same stimuli set were included so the ratio of error to no error sentences was equal. Errors occurred on the last word of the spoken sentences. Written and spoken sentences were presented in blocks and counterbalanced, so that participants either heard the spoken sentences first, or read the written sentences first. Within these blocks, sentences were randomised for each participant. All stimuli were presented using Matlab (version 2016b; Mathworks) and Psychtoolbox (Brainard, 1997).

Procedure. Participants signed the information and consent form, and filled out a music education and preference questionnaire that included demographic and language background questions. They were then informed that they would be reading and listening to sentences that may or may not contain errors, and that these errors could either be meaning errors (semantic), or grammatical errors (syntactic).

Participants practiced with written and spoken sentences that included semantic and syntactic errors, and it was confirmed that they understood the task. In the experiment proper, participants were presented with a sentence, and were asked to indicate on the keyboard whether there was an error (press z) or if there was not an error (press m). Written sentences were presented for 5 seconds on the screen, and spoken sentences were all approximately 2.5 seconds long. Participants wore headphones for the spoken sentences. After completing the written and spoken sentence judgements, participants completed the letter-number sequencing sub-test of the WAIS, as in Experiments 1 and 2. The whole process lasted approximately 30 minutes.

Analysis. Accuracy scores and d prime (d') sensitivity scores were calculated for error detection in each condition for the written and spoken sentences separately (Stanislaw & Todorov, 1999). D prime scores were calculated by subtracting the z score of the false alarm rate from the z score of the hit rate. The semantic and syntactic conditions therefore provided the hit rates for each condition, as there was always an error, and the normal conditions provided the false alarm rate—when participants indicated an error when there was not one. As the spoken sentences had two no error conditions, the average false alarm z scores were calculated for each condition, and then averaged to provide one false alarm z score that was subtracted from the hit z score.

Results

A 2 (sentence type: written, spoken) by 2 (error type: semantic error, syntactic error) repeated measures ANOVA was conducted on the d' scores. This analysis revealed a main effect of sentence type, $F(1, 15) = 20.03$, $p < .001$, $\eta^2 = .57$, a main effect of error type, $F(1, 15) = 14.06$, $p = .002$, $\eta^2 = .48$, and no interaction between sentence type and error type, $F(1, 15) = .32$, $p = .58$. Paired-samples *t*-tests with Holm-Bonferroni corrections for two comparisons showed no difference between semantic (*M*

= 3.77, *SD* = .72), and syntactic (*M* = 4.15, *SD* = .69) error detection in the spoken sentence condition, $t(15) = 1.72$, $p' = .11$, but a significant difference between semantic (*M* = 2.82, *SD* = .71) and syntactic (*M* = 3.39, *SD* = .54) error detection in the written sentence condition, $t(15) = 2.8$, $p' = .026$, $d = .71$.

When the covariate of WMC was added to the same repeated measures ANOVA, there was still a main effect of sentence type, $F(1, 14) = 10.51$, $p = .006$, $\eta^2 = .43$; however, the main effect of error type was no longer significant, $F(1, 14) = 1.69$, $p = .22$, suggesting that WMC can account for differences in error detection between conditions (as observed in Experiment 2). There was no sentence type by WMC interaction $F(1, 14) = 3.4$, $p = .09$, error type by WMC interaction, $F(1, 14) = .04$, $p = .85$, sentence type by error type interaction, $F(1, 14) = .39$, $p = .54$, or three way interaction, $F(1, 14) = .23$, $p = .64$. There was also no between-subjects effect of WMC, $F(1, 14) = .26$, $p = .62$, suggesting that in a single-task situation, participants do not need to rely on WMC to determine whether a sentence has an error or not.

Discussion

The results from Experiment 3 showed that semantic errors were detected worse than syntactic errors in the written sentence condition, but not in the spoken sentence condition. This difference is most likely related to the stimuli used. The written stimuli were from Osterhout and Nicol (1999), and were used because they had previously been shown to elicit the appropriate responses in the brain to semantic and syntactic errors respectively. As such, they were not picked deliberately to include non-ambiguous semantic errors, as it was assumed that errors were behaviourally detectable considering their prior usage. However, the spoken sentences were from Sun et al. (2015), and the most unambiguous semantic errors in the stimuli were deliberately chosen for the semantic error condition, based on the differential performance in

Experiment 1. As such, these errors were likely to be less ambiguous than the errors in the written sentence condition. These results suggest two things. First, if not carefully controlled, semantic errors can be less salient and detectable than syntactic errors, even in a single-task paradigm. This result raises questions about the use of semantic errors as a comparison for syntactic errors in music interference paradigms. Second, the finding that detection of semantic errors was worse than detection of syntactic errors for the spoken sentences in a dual-task situation, but not in a single-task situation, suggests that the dual-task in Experiment 2 interfered with participants' ability to detect semantic errors.

General Discussion

The experiments described in Chapter 5 were designed to investigate whether syntactic interference effects could be observed in the opposite direction to what is usually investigated; that is, from language to music. In Experiment 1, we examined performance on a musical memory (same-different) task as participants were concurrently *reading* sentences with (a) no error, (b) a semantic error, or (c) a syntactic error. In Experiment 2, we examined performance on a melody recall task when participants were *listening* only to the melody, or concurrently listening to sentences that had (a) no error, (b) a semantic error, or (c) a syntactic error. In Experiment 3, the written and spoken sentence stimuli from Experiments 1 and 2 were tested in a single-task paradigm. From this series of experiments, it was revealed that the language syntax manipulations used did not interfere with concurrent processing of the primary music tasks in either Experiment 1 or 2. However, there were a number of interesting results when WMC was taken into account, suggesting that WMC could account for decreased melody recall when sentences were concurrently being processed in Experiment 2, and WMC provided an insight into potential task switching in Experiment 1. Further, semantic errors in sentences were detected worse than syntactic errors in sentences in both Experiments 1 and 2. Experiment 3 revealed that this difference was inherent to the stimuli in Experiment 1, but that the difference was related to the dual-task situation in Experiment 2. This finding suggests that semantic and syntactic errors place different demands on processing resources in dual-task paradigms. Furthermore, a close examination of the melody recall experiment (Experiment 2) revealed that melody recall for the final two notes in the melody was affected differently depending on concurrent sentence condition. The lack of syntactic interference observed from language to music and the nature of dual-task paradigms will be discussed.

Lack of Syntactic Interference From Language to Music

Surprisingly, syntactic interference from language to music was not observed for either of the two musical memory tasks adopted in Experiments 1 and 2: the melodic same-different task from Experiment 1, or the melody recall task from Experiment 2. This lack of interference was unexpected given that interference between music and language should be symmetric based on theories of shared syntactic processing (Koelsch, 2013; Patel, 2008). Interference from language to music has been observed in two recent behavioural experiments (Kunert et al., 2016; Van de Cavey et al., 2017), and a language syntax violation has previously been shown to reduce the neural response (ERAN) to an out-of-key chord (Steinbeis & Koelsch, 2008), suggesting that language to music syntactic interference does occur. The difference between these studies and the current experiments might be explained by methodological differences between experiments in terms of stimuli, the tasks used, and stimulus presentation.

Stimuli. The music stimuli used in Experiments 1 and 2 could be considered typical of experiments that have found syntactic interference effects between music and language. The melodies from Experiment 1 were created from the top line of the chord sequences in Slevc et al. (2009) that were in the style of Bach chorales and strongly tonal. The difference between the Slevc et al. (2009) stimuli and the stimuli in Experiment 1 is that the current stimuli were melodies that contained no out-of-key elements, suggesting that out-of-key elements might be important to produce interference. In Experiment 2, participants were presented with six-note melodies that were to be recalled. One criticism of the six-note melody stimuli could be that the three unique notes within the melodies were too simple to engage syntactic processing resources. For experimental reasons, the melodies had to be simple enough to be recalled. However, the six-note melodies were designed to have syntactic qualities. The three notes chosen were all in the C major key: C (the tonic), G (the fifth), and D (the second). As such, the melodies were strongly tonal. Before each trial, participants heard a C chord and the three notes in order: C, G, D. This sequence provided a tonal context for the listener, and placed each note within this context. To correctly recall the melody, participants had to first identify the correct starting pitch, relate the following pitches to the first pitch, and encode this order. The notes therefore had dependencies between elements that were mapped out within a pitch space. These qualities are syntactic. The reason that interference was not observed in Experiments 1 or 2 may be because integration was not directly manipulated (as in Kunert et al., 2016; Van de Cavey et al., 2017), the melodic stimuli were not complex enough to show interference, or the particular language errors did not draw on the same resources as those required to complete the music task.

The sentence stimuli presented in the current experiments also differed from the stimuli in Kunert et al. (2016) and Van de Cavey et al. (2017). Experiments 1 and 2 described in the current chapter included stimuli with syntactic and semantic errors.

This design is entirely comparable with previous research in which reliable evidence for syntactic interference has been observed (e.g., Hoch et al., 2011; Koelsch et al., 2005). It is also comparable with research showing convergent brain responses to syntactic violations and syntactic garden path sentences (Hopf, Bader, Meng, & Bayer, 2003). Both Kunert et al. (2016) and Van de Cavey et al. (2017) presented garden path sentences that had a disambiguation point, and directly manipulated musical integration at this point. Van de Cavey et al. (2017) suggested that language syntax violations cannot be integrated into the context of the sentence, and therefore should not interfere with the integration of a musical element into a musical context. This suggestion is supported by their findings and somewhat by the findings from Perruchet and Poulin-Charronnat (2013), who found interference with semantic garden path sentences but not semantic anomalies. However, Koelsch (2013) argued that any violation of structure interrupts early syntactic structure building processes, which should interfere with music processing. Therefore, it appears that out-of-key notes interfere with syntactic violations in sentences and in garden path sentences, but that syntactic violations in sentences do not interfere with music processing when there are no out-of-key or complex integrations.

The findings from Van de Cavey et al. (2017) and Perruchet and Poulin-Charronnat (2013) suggest a distinction between syntactic and semantic violations compared to sentences requiring a reanalysis and revision, such as garden path sentences. However, it is possible that there is also a distinction between semantic and syntactic violations in terms of reanalysis. Experiment 2 provided some evidence to suggest that semantic errors resulted in a reanalysis of the sentence, whereas syntactic violations did not. An analysis of recall for the final two notes in the melody revealed that recall increased for the syntactic error and melody only conditions (reminscent of a recency effect; Roberts, 1986), but decreased from the penultimate to final position for the semantic error and no error conditions. This difference suggests that different types of processing are engaged for the two types of errors, likely due to the nature of the violations. The syntactic violations in Experiment 2 were number agreement errors (e.g., *one bags*, *two bag*), that were immediately noticeable, and did not require revision of the sentence to detect an error. On the other hand, both semantic errors and no errors required participants to analyse the sentence for meaning, thereby integrating the final word into the sentence context. Though the sentence errors did not result in a decrease for overall melody recall compared to when there was no error, a very local effect of sentence error on last note recall was observed when the melodies were paired with semantic errors. As such, it appears that the difference between when interference occurs and when it does not could be related to whether reanalysis occurs. Recent research has suggested that for both syntactic and semantic errors, participants could either engage in reanalysis (also leading to better comprehension), or *good-enough* processing that did not result in reanalysis (Metzner, von der Malsburg, Vasishth, & Rösler, 2017). Such findings suggest that the distinction between syntactic and semantic errors is not as clear as might be expected.

Task and stimulus presentation. In the current experiments, syntactic interference was not observed in the primary music task for either a melodic same-different task or a melody recall task. The same-different task could be considered quite similar to the probe tone task in Van de Cavey et al. (2017), as it involved comparing an initial sequence to a secondary exemplar. Van de Cavey et al. (2017) found interference on the music task with the concurrent presentation of garden path sentences, but not sentences with a syntactic error. The combination of the Van de Cavey et al. (2017) study and the current experiments suggests that music processing is not affected by language that contains syntactic violations. This conclusion may suggest that syntactic interference operates asymmetrically between music and language, and an out-of-key note is not equivalent to a syntactic error in language.

In terms of stimulus presentation, previous experiments that found interference from language to music presented language stimuli in non-naturalistic reading settings (word-by-word or section-by-section) with a concurrent note or chord (Kunert et al., 2016; Steinbeis & Koelsch, 2008; Van de Cavey et al., 2017). In Experiments 1 and 2 in this chapter, sentence stimuli were presented either in full on the screen (as would be typically viewed while reading), or sentences were spoken. It is possible that specific alignment of words and chords is necessary to elicit syntactic interference in the current types of tasks, and interference is only observed in music processing with a concurrent integration manipulation (Kunert et al., 2016; Van de Cavey et al., 2017) or concurrent syntactic violations in both music and language (Steinbeis & Koelsch, 2008).

Dual-Task Effects

The music and language syntactic interference literature has a surprising lack of dual-task measurements. Experiments focusing on the effect of music on language processing often have music as a to-be-ignored stimulus, with no music task (e.g., Fedorenko et al., 2009; Fiveash & Pammer, 2014; Hoch et al., 2011; Koelsch et al., 2005; Slevc et al., 2009). When there is a music task, it is often just a detection task (e.g., detect a change in timbre) to ensure participants are paying attention to the stimuli (e.g., Steinbeis & Koelsch, 2008). Similarly, in the studies measuring the effects of language syntax on music syntax processing, participants were only asked to perform a secondary language task on a subset of trials, and these data were only analysed to ensure task adherence (Kunert et al., 2016; Van de Cavey et al., 2017). Indeed, Kunert et al. (2016) suggested that these data were of “limited interest”. However, the current experiments suggest that performance on the secondary task may provide more insight into the nature of shared processing between music and language than previously thought.

In addition, dual-stimulus paradigms where participants are tested on two tasks at once (whether or not the second task data is analysed) might result in participants using different strategies to process both streams of information at the same time. A close examination of Experiment 1 suggested that participants were task switching to concurrently process the music and language streams. Upon encountering a syntactic error, participants may have quickly decided the error and switched their attention to the music. This suggestion is supported by evidence that WMC was correlated with performance on the music task while reading the no error and semantic error sentences, but not the syntactic error sentences, and that performance was worse on last word judgements in the syntactic condition compared to the other two conditions. Previous research has also shown that syntactic errors resulted in a rapid change in eye-movements, but a quick recovery back to baseline reading, whereas semantic errors led to a more lasting effect on eye-movements, and increased regressions (looking back to the beginning of the sentence) in eye-movements (Braze et al., 2002). It is possible that these effects were also occurring in Experiment 1. The results of Experiment 1 therefore suggest that dual-task paradigms where syntactic errors are not aligned in the two streams (or there are no errors) may lead to different effects than in experiments that align the errors in both domains and only have one task. This result is important for generalising music and language syntax interference studies into real world situations.

Experiment 2 revealed that WMC could account for a number of differences between conditions. In Experiment 2, melodies presented with sentences were recalled significantly worse than melodies presented without sentences. This finding is not surprising considering a dual-task situation would place a greater burden on processing resources. However, when WMC was taken into account, there was no difference in melody recall between the melody only single-task, and the melody and sentence dual-task, suggesting that WMC accounted for the effects of the dual-task on melody recall. Furthermore, Experiment 2 showed that participants were significantly worse at detecting errors in the semantic error sentences compared to the syntactic error sentences. This difference also appeared related to WMC, and disappeared when WMC was taken into account. Confirming this finding, Experiment 3 showed that there was no difference in semantic and syntactic error detection in a single-task with the same stimuli from Experiment 2, suggesting that the dual-task decreased detection of semantic errors. This finding suggests that it is important to measure performance on both tasks in a dual-task situation, and that WMC plays a large role in interference in a dual-task situation.

Conclusion

The current experiments raise questions about the conditions under which syntactic interference can be observed between music and language, and the importance of measuring performance on both tasks in dual-task experiments. Interference was not observed from language to music in either a melodic same-different task, or a melody recall task, suggesting that syntactic errors in sentences do not interfere with melodic processing in these two tasks. However, WMC appears to be intricately tied to the processing of two simultaneous tasks in a dual-task situation. Future research should systematically test when interference does and does not occur, manipulating the type of sentences and music tasks used, and incorporating measures of WMC. The results presented in this chapter also suggest that it may be problematic to compare semantic and syntactic errors, because (a) in some stimuli semantic errors are detected worse than syntactic errors, and (b) in a dual-task situation, semantic errors are not detected as well as syntactic errors. The current experiments therefore suggest that syntactic interference is a very specific effect that can only be observed behaviourally under specific conditions.

Acknowledgments. I would like to thank Glenn Schellenberg for input on the design of Experiment 1, and for access to facilities and resources at University of Toronto, Mississauga for testing participants. Thank you to Shayan Alam for help running participants for Experiment 1.

References

Baddeley, A. (2000). The episodic buffer: A new component of working memory. *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A., & Hitch, G. (1974). Working memory. In H. B. Gordon (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31*(1), 25-44. doi:10.1023/a:1014324220455

Burunat, I., Alluri, V., Toiviainen, P., Numminen, J., & Brattico, E. (2014). Dynamics of brain activity underlying working memory for music in a naturalistic condition. *Cortex, 57*, 254-269. doi: 10.1016/j.cortex.2014.04.012

Crowder, R. G., & Greene, R. L. (2000). Serial learning: Cognition and behavior. In E. Tulving & F. I. Craik (Eds.), *Oxford handbook of memory*. New York: Oxford University Press.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition, 37*(1), 1-9. doi:10.3758/MC.37.1.1

Fiebach, C., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during pasing: The processing of German WH-questions. *Journal of Memory and Language, 47*, 250-272.

Fiveash, A., McArthur, G., & Thompson, W. F. (under review). Music and language: Syntactic interference without syntactic violations.

Chapter 5: Language to Music Interference

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music, 42*(2), 190-209. doi:10.1177/0305735612463949

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84. doi:10.1016/S1364-6613(00)01839-8

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 351-360. doi:10.1016/S0022-5371(66)80044-0

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory, 13*(3-4), 318-324. doi:10.1080/09658210344000350

Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: Syntactic and semantic structures. *Frontiers in Psychology, 2*, 112. doi:10.3389/fpsyg.2011.00112

Hopf, J., Bader, M., Meng, M., & Bayer, J. (2003). Is human sentence parsing serial or parallel? Evidence from event-related brain potentials. *Cognitive Brain Research, 15*(2), 165-177. doi:10.1016/S0926-6410(02)00149-0

Jahnke, J. C. (1963). Serial position effects in immediate serial recall. *Journal of Verbal Learning and Verbal Behavior, 2*(3), 284-287. doi:10.1016/S0022-5371(63)80095-X

Jahnke, J. C. (1965). Primacy and recency effects in serial-position curves of immediate recall. *Journal of Experimental Psychology, 70*(1), 130-132.

King, J., & Just, M. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*(5), 580-602.

Kljajević, V. (2010). Is syntactic working memory language specific? *Psihologija, 43*(1), 85-101. doi: 10.2298/psi1001085k

Chapter 5: Language to Music Interference

Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology, 15*(2), 207-212. doi:10.1016/j.conb.2005.03.005

Koelsch, S. (2013). *Brain and music*. Oxford, UK: John Wiley & Sons.

Koelsch, S., Gunter, T., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and music: An ERP study. *Journal of Cognitive Neuroscience, 17*(10), 1565-1577.

Kunert, R., Willems, R. M., & Hagoort, P. (2016). Language influences music harmony perception: Effects of shared syntactic integration resources beyond attention. *Royal Society Open Science, 3*(2). doi:10.1098/rsos.150685

Mathworks (R2016b). Matlab. Massachusetts, United States: Mathworks.

Metzner, P., von der Malsburg, T., Vasishth, S., & Rösler, F. (2017). The importance of reading naturally: Evidence from combined recordings of eye movements and electric brain potentials. *Cognitive Science, 41*, 1232-1263. doi:10.1111/cogs.12384

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*(3), 283-317. doi:10.1080/016909699386310

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674-681.

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

Patston, L. M., & Tippett, L. J. (2011). The effect of background music on cognitive performance in musicians and nonmusicians. *Music Perception, 29*(2), 173-183. doi: 10.1525/mp.2011.29.2.173

Chapter 5: Language to Music Interference

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin & Review, 20*(2), 310-317. doi:10.3758/s13423-012-0344-5

Roberts, L. A. (1986). Modality and suffix effects in memory for melodic and harmonic musical materials. *Cognitive Psychology, 18*(2), 123-157. doi:10.1016/0010-0285(86)90010-1

Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior, 9*(1), 99-105. doi:10.1016/S0022-5371(70)80015-9

Slevc, L. R., & Okada, B. M. (2014). Processing structure in language and music: A case for shared reliance on cognitive control. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-014-0712-4

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review, 16*(2), 374-381. doi:10.3758/16.2.374

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18*(5), 1169-1178. doi:10.1093/cercor/bhm149

Sun, Y., Lu, X., Ho, H. T., Johnson, B., & Thompson, W. F. (2015). Exploring music-syntactic processing and language-syntactic processing in congenital amusia using MEG and EEG. In J. Ginsborg, A. Lamont, M. Phillips, & S. Bramley (Eds.), *Proceedings of the 9th triennial conference of the European society for* *the cognitive sciences of music (ESCOM)* (pp. 765-770). Manchester, UK: Royal Nothern College of Music.

Tillmann, B., & Bigand, E. (2015). A commentary on "A commentary on: 'Neural overlap in processing music and speech'". *Frontiers in Human Neuroscience, 9*. doi:10.3389/fnhum.2015.00491

Van de Cavey, J., Severens, E., & Hartsuiker, R. J. (2017). Shared structuring resources across domains: Double task effects from linguistic processing on the structural integration of pitch sequences. *The Quarterly Journal of Experimental Psychology, 70*(8), 1633-1645. doi:10.1080/17470218.2016.1195852

Vos, S. H., Gunter, T. C., Kolk, H. H., & Mulder, G. (2001) Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology, 38*(1), 41-63.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - Fourth Edition*. San Antonio, TX: Pearson.

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. doi:10.3758/MC.38.2.163

Chapter 6

Discussion

Chapter 6: Discussion

This thesis aimed to elucidate the nature of syntactic processing in music and language. My objective was to investigate the conditions under which syntactic interference occurs, in line with the two major theories of shared syntactic processing between music and language—the shared syntactic integration resource hypothesis (SSIRH; Patel, 2008), and the syntactic equivalence hypothesis (SEH; Koelsch, 2013). Research supporting these theories is largely based on stimuli involving salient disruptions to the auditory stream and to syntactic structure, and largely examines the effects of music processing on language processing tasks. An assessment of the research in this area revealed that we do not yet have a strong understanding of the effect of music syntax violations on the syntactic representation developed through auditory streaming, or whether violations of language syntax affect music processing in the same way that violations of music syntax affect language processing. Furthermore, studies measuring the effect of music on language processing are unable to provide insights into the concurrent processing of music and language in a dual-stimulus situation, or whether a cognitive trade-off occurs with the simultaneous processing of music and language.

To elucidate the nature of syntactic processing in music and language and to address these limitations in the literature, this research pursued three related lines of enquiry. First, I investigated whether syntactic interference could be observed in stimuli without violations of syntax. In Chapter 2, I established that syntactic interference does not depend on specific processes engaged with syntactic violations in language and music. However, it does appear to depend on a sensitive dependent variable such as recall for sentences, as opposed to a less sensitive variable such as language comprehension. Having established that syntactic interference between music and language can be induced using stimuli without violations in either domain, the focus of the thesis shifted to the role of auditory streaming in syntactic processing. In Chapters 2 and 3, I found that disrupting a musical auditory stream altered the way that music syntax was processed in the brain. I then investigated whether syntactic violations operate symmetrically for music and language, or if music syntax interrupts language processing more than language syntax interrupts music processing. In Chapter 5, I did not observe syntactic interference from language to music in two experiments that used different stimuli and tasks. However, these experiments did reveal a number of other interesting effects, such as position effects (i.e., semantic and syntactic errors in sentences had different effects on recall for the final two notes in a musical sequence) and working memory capacity (WMC) effects (i.e., WMC can account for many of the differences between conditions, suggesting it is inherently linked to syntactic processing) that may inform theories of concurrent music and language processing.

The experiments presented in this thesis suggest that interference between music and language is a specific effect that is only observable at the behavioural level under very precise conditions. In addition to addressing the three questions above, the research in this thesis has shed light on the nature of dual-stimulus processing and the importance of measuring performance on both music and language tasks. Further, it has explored the validity of comparing semantic and syntactic errors in language, the importance of equating experimental stimuli for distraction, and the necessity of having a sensitive dependent variable (such as sentence recall) in experiments investigating music and language syntactic interference. In this final chapter, I first outline the findings presented in the thesis in relation to the three main research areas: syntactic interference without violations, auditory streaming and syntactic processing, and syntactic interference from music to language. I then discuss these findings in relation to the SSIRH and the SEH. Following this, I propose a new *competitive attention and* *prioritisation model* (CAP) to account for the pattern of results in the current thesis, as well as previous investigations of music and language processing. The CAP model places syntactic processing within a larger cognitive framework, and emphasises processes of task prioritisation, field of attention, and auditory streaming. The final section is a discussion of challenges and prospects for the CAP model, and suggestions for future research based on this model.

Syntactic Interference Without Violations

Most previous studies investigating syntactic processing in music and language, including studies of syntactic interference, have used stimuli with syntactic violations. For music, a syntactic violation typically involves inserting out-of-key notes or chords into an otherwise conventional sequence (Fedorenko, Patel, Casasanto, Winawer, & Gibson, 2009; Fiveash & Pammer, 2014; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Slevc, Rosenberg, & Patel, 2009). For language, a syntactic violation usually involves inserting a grammatical error within a sentence (Hoch, Poulin-Charronnat, & Tillmann, 2011; Koelsch et al., 2005; Steinbeis & Koelsch, 2008). When such violations occur in music and language simultaneously, interference is often observed. However, it is unclear whether the interference is specific to syntactic processing, or whether it reflects more general processes of error detection or distraction that are specific to the violation itself (Tillmann & Bigand, 2015). In Chapters 2 and 4, I investigated this uncertainty using stimuli that did not contain musical or grammatical violations.

To investigate syntactic interference in music and language without violations, Chapter 2, Experiments 1 and 3, involved participants listening to melodies or environmental sounds while concurrently reading and then recalling sentences and word-lists. Language recall was therefore the primary task, and there was no secondary music task. In Experiment 1, participants made more errors in sentence recall whilst concurrently listening to normal melodies than when listening to environmental sounds, as predicted. There was no difference in word-list recall, suggesting that listening to melodies specifically interfered with the processing of language syntax. Experiment 1 therefore established that music interferes with language processing in the absence of violations of syntactic structure.

In Chapter 2, Experiment 3, I investigated the processing of music and language without syntactic violations while participants were attending either to a melodic auditory stream (syntactic) or an environmental auditory stream (non-syntactic). Participants listened to stimuli with melodies presented to one ear, and environmental sound chunks (a different environmental sound each second) presented to the other ear. The participants' attention was directed to either the melodies or the environmental sounds by asking them to indicate if they heard a *target* (an out-of-key note in the melodies, or a gong sound in the environmental sounds). This design ensured that participants always received an auditory signal that included both a melodic stream and an environmental sound stream—the only difference between conditions was where they directed their attention. The target trials (including an out-of-key note or a gong) were not included in the main language recall analysis, and were only included to direct attention. Therefore, the study permitted an assessment of interference effects for stimuli that did not include syntactic violations. While directing attention to one stream, participants read a word-list, a basic sentence, or a complex sentence. Basic and complex versions of sentence stimuli were included to determine whether there was increased syntactic interference with more complex sentences. After each stimulus, participants recalled the word-list or sentence aloud as per Experiment 1. They then indicated whether they noticed the target. The primary task was therefore language recall, and the secondary task was target detection in the auditory stimuli.

Chapter 2, Experiment 3 revealed that participants made more errors in sentence recall (but not word-list recall) when attention was directed towards the melodies compared to environmental sounds. This finding is interesting considering that the auditory stimuli always contained both environmental sounds and melodies—the only difference was where attention was directed. There was no reliable difference between errors in basic and complex sentence recall; however, recall for complex sentences seemed to be the most affected by attention to melodies. A further analysis of target detection accuracy revealed that the detection of out-of-key notes was significantly reduced when participants were reading sentences compared to word-lists. This effect did not occur for the detection of gongs, suggesting that the syntactic structure in the sentences interfered with the detection of out-of-key notes in the secondary task of target detection. Experiment 3 therefore revealed syntactic interference on the secondary task of target detection as well as the primary task of language recall. Chapter 2, Experiments 1 and 3 therefore revealed that with a sensitive dependent variable such as errors in sentence recall, syntactic interference could be observed with stimuli that did not violate syntactic structure.

In Chapter 4, Experiments 1 and 2, I aimed to separate syntax-specific processing from more general *complexity* processing, again, using stimuli without violations. This design was motivated by previous studies which have suggested that increased complexity in sentences also introduces additional confounds such as increased working memory load (e.g., Fiebach, Schlesewsky, & Friederici, 2002) that might be related to more general processes of interference rather than specific syntactic interference. To separate complexity from syntactic processing, I manipulated music and environmental sounds to create basic and complex versions of stimuli. Increased complexity in music introduced syntactic structure, whereas increased complexity in environmental sounds did not. In Chapter 4, Experiments 1 and 2, participants listened to the auditory stimuli whilst engaged in either a language comprehension task (reading a complex sentence and then answering a comprehension question), or a visuospatial search task (searching two images to detect a difference) as primary tasks. In Chapter 4, Experiment 1, participants were also asked to decide if the auditory stimulus was musical or environmental (secondary task). In Chapter 4, Experiment 2, participants were also asked to decide of the auditory stimulus was basic or complex—a more complex judgement. Neither experiment showed that the primary language comprehension task or the primary visuospatial search task were affected by the concurrent presentation of basic or complex musical or environmental stimuli. This finding is surprising, considering the large differences in auditory stimulus type and complexity. It therefore appears that language comprehension and visuospatial processing are relatively robust to background auditory stimuli.

Although performance on the primary language comprehension task did not differ depending on the concurrent auditory condition, Chapter 4, Experiment 2 revealed interference in the secondary task involving auditory complexity judgements. While engaged in the visuospatial search task, participants were significantly better at judging complexity in music than judging complexity in environmental sounds, suggesting that judging complexity in music was an easier task. However, there was no difference between judging musical and environmental complexity when participants were engaged in the language comprehension task, suggesting that the language task interfered with judgements of musical complexity. This suggestion is supported by the finding that participants performed significantly worse on the musical complexity judgements while engaged in the language task than while engaged in the visuospatial search task, but there was no difference in environmental complexity judgements depending on the concurrent task. Therefore, although interference was not observed for the primary task of language comprehension, interference was observed on the secondary task of complexity judgements. Interestingly, these effects disappeared when WMC was taken into account, suggesting that WMC plays a crucial role in syntactic processing.

The four experiments described above were designed to investigate whether syntactic interference could be observed in the absence of explicit violations of syntactic structure. This investigation revealed that interference can be observed with a sensitive dependent variable (e.g., recall for language), and in a secondary music task (e.g., detection of out-of-key notes, musical complexity judgements). With the current stimuli and tasks, manipulations of syntax or complexity do not appear to affect language comprehension or performance on a visuospatial search task. These findings suggest that syntactic interference from music to language is very specific, and may require alignment of difficult syntactic processing requirements in both domains simultaneously in order to elicit interference for measures such as language comprehension (as in Fedorenko et al., 2009). However, syntactic violations or alignment of difficult integration were not necessary to elicit interference with a sensitive dependent variable such as sentence recall. Furthermore, interference in sentence recall was differentially affected by where participants were directing their attention, suggesting that participants have some control over how much a syntactic musical sequence interferes with memory for sentences.

Auditory Streaming and Syntactic Processing

For syntactic processing to occur, it has been suggested that incoming elements must first be organised into a coherent stream of related information (Koelsch, 2013). For auditory stimuli, auditory scene analysis processes incoming information into auditory streams via Gestalt grouping principles (Bregman, 1990). For example, notes that are close together in pitch or time (Gestalt principle of proximity) are likely to be grouped together, as are notes with the same timbre (Gestalt principle of similarity). Therefore, there is a strong connection between auditory streaming and syntactic processing, as sequences that are not processed as a coherent auditory stream are unlikely to be processed syntactically. Whether auditory streaming is a prerequisite for syntactic processing, or whether the two processes interact with one another in complex ways, is not fully understood. However, it is clear that a full understanding of syntactic processing implicates processes related to auditory scene analysis.

In Chapter 2, Experiment 1, melodies with alternating timbres were initially included to match the level of distraction of the scrambled stimuli. Based on the changing-state theory (Hughes, Vachon, & Jones, 2007; Jones, Hughes, & Macken, 2010), I predicted that the changes in timbre would result in an additive effect of syntactic interference and distraction, as sequences that change state (e.g., change pitch, change timbre) should be more distracting than a steady-state sound (a sound that does not change state). Instead, the three-timbre melodies resulted in fewer errors on the sentence recall task than melodies played with one timbre. It was hypothesised that the alternating timbres disrupted auditory streaming processes (Bregman, 1990), resulting in a less coherent syntactic representation of the melody, and therefore less interference with language syntax. In Chapter 2, Experiment 2, and in Chapter 3, I investigated the hypothesis that alternating timbres interrupted auditory streaming processes and resulted in reduced syntactic processing. In Chapter 2, Experiment 2, this hypothesis was first investigated behaviourally. In Chapter 3, this hypothesis was investigated with an event-related potential (ERP) study that measured the electrophysiological processing of music and language syntax under one- and three-timbre conditions.

Chapter 2, Experiment 2 was a behavioural same-different task, where participants indicated whether two melodies were the same or whether they differed by one note. Participants were significantly worse at this task when both of the melodies were played with three timbres compared to when both melodies were played with one timbre. Poorer performance in the three-timbre condition suggested that participants were less sensitive to the syntax in melodies with alternating timbres.

To investigate whether alternating timbres influence how the brain responds to a syntactic sequence, and to ensure the interpretation in Chapter 2 was correct, an ERP study was conducted. In this study, reported in Chapter 3, I investigated how the brain responds to syntactic violations depending on whether the melody is played with one timbre or three timbres. Extending this paradigm to language processing, I also created stimuli to investigate the brain response to phrase-structure violations in sentences depending on whether the sentences were spoken by one speaker or by three speakers. I predicted that the brain response to syntactic violations would be significantly reduced in the three-timbre conditions compared to the one-timbre conditions.

Syntactic violations in music and language elicit specific ERP components that can be used as an index of syntactic processing. Out-of-key elements in music elicit the early right anterior negativity (ERAN; Koelsch, 2013; Koelsch et al., 2001) and the P600 (Patel et al., 1998). Grammatical errors in language can elicit different components depending on the type of violation; however, these include the early left anterior negativity (ELAN), the left anterior negativity (LAN) and the P600 (Friederici, 2002). I focused on the early brain responses to violations—the ERAN in music, and the ELAN and LAN in language. Both the one-timbre and three-timbre sequences presented in the ERP study in Chapter 3 elicited the expected components. Melodies with out-of-key notes elicited an ERAN, and sentences with phrase-structure violations elicited a LAN, as predicted. There was also a small ELAN for the one-timbre sentence stimuli, but not the three-timbre sentence stimuli, partially supporting the hypothesis of reduced syntactic processing in the three-timbre condition for sentence stimuli. The main finding presented in Chapter 3 was that the ERAN response to an out-of-key note in a melody was significantly reduced when the melody contained three-timbres compared to one-timbre. The reduction of the ERAN in the three-timbre condition supports the hypothesis that the brain does not process syntax to the same extent when the auditory stream is disrupted by alternating timbres. The sentence stimuli showed a similar pattern—the LAN was visibly reduced in the three-timbre condition; however, this reduction did not differ statistically. The elimination of the ELAN for the three-timbre sentence condition provides some evidence for a difference between one- and three-timbre sentence stimuli at early stages; however, this reduction is clearly observed in the difference between one- and three-timbre melodic stimuli.

The aforementioned experiments in Chapter 2 and Chapter 3 suggest that interrupting an auditory stream with alternating timbres affects behavioural and electrophysiological measures of syntactic processing in music. The ERP study was designed to test whether auditory streaming and syntactic processing operate in parallel across music and language. There was evidence to suggest that the ELAN was elicited in the one-timbre sentence stimuli but not the three-timbre sentence stimuli. The LAN was also visibly reduced in the three-timbre sentence condition compared to the one-timbre sentence condition. As this reduction was not significant, strong parallels cannot be drawn between music and language processing in relation to auditory streaming and syntactic processing. However, the findings from Chapter 2, Experiment 2, and Chapter 3 revealed for the first time that interrupting an auditory stream with alternating timbres reduced the cognitive response to syntactic violations in music. Future research should continue to investigate this claim, as similar findings would be expected for any type of manipulation that interrupted auditory streaming, such as changes in location of sound, increased distance between pitches, or disrupted rhythm.

In addition to the three-timbre results, Chapter 2, Experiment 1 revealed that scrambled melodies resulted in the most errors in recall for both sentences and word-lists, possibly because the scrambled melodies were more distracting (as they also affected word-list recall) and more complex (as assessed by a model of musical complexity; Eerola, 2016) than normal melodies. Thus, scrambled melodies may not be a good measure of non-syntactic processing, though similar scrambled stimuli have been used in the past (Abrams et al., 2011; Fedorenko, Behr, & Kanwisher, 2011; Levitin & Menon, 2003; Tillmann & Bigand, 2001). The difference between the scrambled stimuli used in Experiment 1 and the scrambled stimuli in previous experiments is that individual notes were scrambled in Experiment 1, whereas in previous studies, music was segmented into small sections (e.g., 250 milliseconds), and these sections were scrambled (e.g., Abrams et al., 2011; Levitin & Menon, 2003). It is likely that the scrambled stimuli in Experiment 1 were perceived as one auditory stream, because of the proximal pitches, proximal note distances, and notes of the same timbre. Therefore, by scrambling the notes within an intact auditory stream, a greater syntactic cost may have been imposed as the brain tried to make sense of the unpredictable syntactic sequence. However, scrambling music based on the segmentation of small sections (as in previous scrambled stimuli) is likely to interrupt the auditory stream, and eliminate (or reduce) syntactic processing. These results point to a central role of auditory streaming in syntactic processing.

Syntactic Interference From Language to Music

Returning to interference paradigms, Chapter 5, Experiments 1 and 2 were designed to investigate whether interference could be observed from language to music when music processing was the primary task (i.e., a melodic same-different task and a melody recall task) and language processing was the secondary task (i.e., syntactic and semantic error detection in language). This reversal of the usual measurement of music syntax on language syntax processing allowed much-needed insight into whether music-language syntactic interference is symmetrical (e.g., occurs in both directions), or whether language processing is particularly susceptible to interference by music. The majority of experiments investigating music and language syntactic interference have investigated the effect of music on language processing (exceptions: Kunert, Willems, & Hagoort, 2016; Van de Cavey, Severens, & Hartsuiker, 2017), and it is unclear whether syntactic violations in language affect music syntax processing in the same way that syntactic violations in music affect language processing. Chapter 5, Experiments 1 and 2 included sentence stimuli with syntactic and semantic errors to investigate whether interference could be observed in music processing when participants were reading or listening to sentences with syntactic errors compared to sentences with semantic errors.

To investigate the effect of language on music processing, in Chapter 5, Experiment 1, participants listened to two melodies and judged whether they were the same or different. While the first melody was playing, participants concurrently read a sentence that contained (a) no error, (b) a semantic error, or (c) a syntactic error. They then heard the second melody unaccompanied by written sentences. Participants first indicated whether the melodies were the same or different, and received feedback on this task, indicating that it was the primary task. Second, they indicated whether there was an error in the sentence, and third, they indicated what the last word in the sentence was to ensure the whole sentence was read. It was surprising to find no difference in melodic same-different judgements depending on language condition because it was predicted that a syntactic error in language should interfere with music processing in the same way that a syntactic error in music interferes with language processing (Fedorenko et al., 2009; Fiveash & Pammer, 2014; Slevc et al., 2009). It is possible that presenting the second melody with no accompanying sentence may have allowed for a recovery of any syntactic interference that may have been present when the first melody and the sentence were concurrently presented. Furthermore, the presentation of the full written sentence with the error located in the middle may have resulted in participants switching their attention to the music as soon as they detected the error. This strategy may have also eliminated any syntactic interference that was present. Chapter 5, Experiment 1 also revealed that participants were significantly worse at detecting semantic errors compared to syntactic errors in sentences.

Chapter 5, Experiment 2 was designed to approach language to music interference in a different way. In this experiment, participants simultaneously heard a six-note melody and a spoken sentence. Six-note melodies were presented with spoken sentences that ended with (a) no error, (b) a semantic error, or (c) a syntactic error. Participants first recalled the melody (the primary task) and then indicated whether there was an error in the spoken sentences (the secondary task). Because the spoken sentences were presented auditorily, participants had to process both streams simultaneously. It was predicted that syntactic errors in spoken sentences would interrupt syntactic structure building processes in these sentences (Koelsch, 2013) and reduce recall for melodies, as the syntactic error in language should also disrupt the syntactic representation of the melody. The results showed that the simultaneous processing of a melody and a sentence reduced performance on the melody recall task compared to only processing a melody (as would be expected with the addition of a second stimulus and task). However, there was no difference between the sentence conditions when examining total melody recall, suggesting that the type of sentence error did not affect recall for full melodic sequences. Interestingly, the reduced recall for melodies when they were concurrently presented with spoken sentences could be explained by controlling for differences in WMC. This finding suggests that working memory plays an important role in decreased performance in a dual-task situation, and is intricately tied to syntactic processing in music and language. As in Experiment 1, participants were also worse at detecting semantic errors compared to syntactic errors in the secondary language task.

Immediate serial recall for melodies typically produces a recency effect, whereby recall for notes towards the end of a sequence is superior to recall for notes in the middle of a sequence (Roberts, 1986). An examination of recall for the final two notes in the melodies (where the language errors occurred) from Chapter 5, Experiment 2 revealed that recall for the melody only and syntactic error conditions increased from the penultimate to the final position, as would be expected. However, recall for the semantic error and no error conditions decreased from the penultimate to the final position. This difference indicated that melodic processing may be affected by the processing required for the concurrent sentence condition. It is possible that the concurrent semantic error and no error sentence conditions decreased performance for melody recall in the final position because they required more processing resources to determine whether the final word fitted the context of the sentence. The syntactic errors on the other hand could immediately be determined as errors, releasing resources to recall the melody. This interpretation is supported by an eye-tracking study which showed that semantic errors resulted in more regressive eye-movements and a prolonged effect on eye-movements compared to syntactic errors, which had more immediate and short-term effects (Braze, Shankweiler, Ni, & Palumbo, 2002). The finding that the semantic error and no error sentence conditions eliminated the recency effect for serial recall of melodies suggests that interference may be related to processes of reanalysis and costs of working memory rather than syntactic processing specifically (Slevc & Okada, 2014; Van de Cavey et al., 2017).

Neither Experiment 1 nor 2 in Chapter 5 showed effects of language syntax violations on music processing, either in a same-different task or a melody recall task. These results could have occurred for a number of reasons, including the specific language violations used, the alignment of the two tasks, and the tasks themselves. Shared syntactic processing theories appear to predict symmetry in syntactic processing, but it is possible that violations of language syntax do not affect music syntax processing in the same way that violations of music syntax affect language syntax processing. However, ERP evidence has shown that the ERAN to a music violation is affected by a syntactic violation in language (Steinbeis & Koelsch, 2008), providing some support that syntactic interference operates in both directions. Studies by Kunert et al. (2016) and Van de Cavey et al. (2017) have also observed interference in music processing when integration of elements was directly manipulated, rather than violations specifically. Further, Chapter 2, Experiment 3 showed that detection of out-of-key notes in music was significantly worse when participants were concurrently reading sentences compared to word-lists (an effect that did not occur for gong detection). These results suggest that language processing can affect music processing, and that future research is necessary to elucidate the conditions under which this interference occurs.

Experiments 1 and 2 in Chapter 5 revealed that detection of semantic errors was worse than detection of syntactic errors in sentences. To investigate whether this difference was inherent to the stimuli, or a result of the dual-task, the stimuli from Experiment 1 (written sentences) and Experiment 2 (spoken sentences) were presented to a different group of participants in a single-task paradigm in Experiment 3. Semantic errors were detected less than syntactic errors in the written sentences from Experiment 1, suggesting that the semantic errors in these stimuli were more difficult to detect than syntactic errors. However, there was no difference between semantic and syntactic error detection for the spoken sentences from Experiment 2. This finding suggests that the melody recall task in Experiment 2 interfered with the detection of semantic errors in language. Chapter 5, Experiments 1 and 2 did not show syntactic interference from language to music in the primary tasks. However, they did reveal a number of other interesting results such as differences between semantic and syntactic errors, distinctions between primary and secondary tasks, and different effects of syntactic and semantic errors on recall for the final two notes of the melodic sequence.

Implications for Current Theories of Shared Processing

The experiments presented in this thesis broaden the scope and extend the understanding of shared processing between music and language. The two dominant theories of shared music and language processing resources (the SSIRH and the SEH) are unable to fully account for the pattern of results described throughout this thesis, as they do not consider the nature of dual-stimulus experiments, stimuli without violations of syntax, the importance of attention when processing two streams simultaneously, or the role of working memory in syntactic processing. They are therefore limited in their predictions, and tied to situations with violations of structure (the SEH) or difficult syntactic integrations (the SSIRH). The SSIRH predicts that syntactic interference between music and language will occur when syntactic integration is difficult in both domains simultaneously. It is unclear from existing descriptions of the SSIRH whether this interference extends to naturalistic listening conditions without aligned integrational difficulties or violations of syntax. The SEH is an extension of the SSIRH, as it discusses shared integrational resources and initial structure building resources, and suggests overlap in both aspects of processing. Most experiments supporting the SEH are based on the ERAN component elicited in music with an out-of-key element (e.g., Koelsch, 2013; Koelsch et al., 2005; Steinbeis & Koelsch, 2008). From the SEH, it is also unclear whether there is overlap in processing incoming syntactic information when there are no violations of syntax or unexpected elements. Both models are limited because they do not take into account more naturalistic listening situations without structural violations or difficult syntactic integration, and do not account for the role of working memory in syntactic processing.

Furthermore, although most studies evaluating the SSIRH and the SEH present music and language simultaneously, they usually only include one task, and often involve violations of syntax (e.g., Fedorenko et al., 2009; Fiveash & Pammer, 2014; Hoch et al., 2011; Koelsch et al., 2005; Slevc et al., 2009). In dual-stimulus situations, it appears to be critical to measure processing of both stimulus types to investigate whether there is a cognitive trade-off occurring when processing two streams simultaneously. Furthermore, studies supporting the SSIRH and the SEH tend to manipulate syntactic processing at a very local level (e.g., specific alignment of out-of-key notes with language errors or difficult aspects of a sentence). Such studies are unable to provide insight into the processing of two full syntactic sequences when there are no salient, distracting elements. A more comprehensive model of syntactic processing without violations is therefore necessary and important to understand how the brain processes incoming syntactic information in naturalistic settings.

The results presented in this thesis suggest that although music and language do appear to show interference without violations—which is reflective of shared resources (Chapter 2, Experiments 1 and 3)—this effect is difficult to capture behaviourally (Chapter 4, Experiments 1 and 2), and is not observed in structurally simple music tasks (Chapter 5, Experiments 1 and 2). Therefore, a more comprehensive model of music and language processing appears necessary to account for the range of findings where syntactic overlap between music and language processing is not observed. The current results suggest that participants prioritise the primary stimulus or task, and that interference only occurs under certain conditions, and sometimes only in the secondary task. By not measuring processing of both the music and language stimuli, previous research has been unable to reveal this interplay between task prioritisation, attention, and dual-stimulus processing. Steinbeis and Koelsch (2008) have shown that simultaneous attention to music and language alters both the type of brain response to a syntactic violation (an additional N5 component is present), and interactions with language processing (the ERAN is reduced when paired with a language syntax violation). These findings suggest that task goals (e.g., attend to both domains, attend to one domain) affect how the brain processes incoming information, and that it is important to measure performance on both tasks. The following section presents a model of music and language dual-stimulus processing that aims to describe syntactic processing in music and language when there are no violations of syntax, and participants engage with both tasks.

The Competitive Attention and Prioritisation (CAP) Model

The CAP model was developed to build upon the SSIRH and the SEH by placing syntactic processing within a larger cognitive framework. This framework can account for situations where there are no violations of syntax, and situations where participants are engaged in two tasks simultaneously. The CAP model incorporates auditory streaming theory (Bregman, 1990) and dual-stimulus processing to provide a more comprehensive account of syntactic processing in music and language. The CAP model suggests a dynamic and competitive interplay between task prioritisation (depth of processing) and field of attention (breadth of processing) in a dual-stimulus situation, based on top-down task goals and bottom-up stimulus characteristics (see Figure 1). Bottom-up processing is based on the auditory streaming of incoming information if there are two auditory streams (e.g., music and spoken sentences), or both auditory streaming and visual processing if there are both auditory and visual stimuli (e.g., music and written sentences). The stimulus characteristics and nature of the incoming elements affect how they are processed cognitively. At the same time, processing of incoming information is directly informed by top-down, long-term knowledge representations (e.g., tonal hierarchies in music and knowledge of lexical and syntactic features in language), and task goals of the situation. Task goals can be explicit (e.g., recall language while ignoring music) or implicit (e.g., participant is given feedback on only one task) to an experiment, or can be present in a real-life situation (e.g., read a book while listening to music). The CAP model differs from the SSIRH and the SEH in that it offers an account of naturalistic music and language processing that takes into account dual-stimulus situations, and incorporates working memory and attentional processes. The CAP model is consistent with capacity models and resource models of attention, such as those discussed by Kahneman (1973), Pashler and Johnston (1998), and Wickens (2002). It is also consistent with the idea of shared processing resources in the SSIRH (Patel, 2008).

The proposed limited-capacity processing resources in the CAP model are suggested to involve working memory and attention. The central executive component of the Baddeley and Hitch (1974) working memory model reflects the types of processes suggested in CAP: focusing attention, dividing attention, task switching, and interfacing with long-term memory, in addition to the maintenance of information and the integration of incoming elements into an evolving, coherent representation. As reported in Chapters 4 and 5, WMC is positively correlated with successful performance in dual-task situations, and appears to account for much of the variance related to syntactic interference (e.g., in Chapter 4, Experiment 2). As reported in Chapter 5, differences in WMC revealed different task switching tactics depending on concurrent language condition in a melodic same-different task in Experiment 1. Further, when WMC was controlled for, there was no difference between recall for a melody presented alone compared to recall for a melody presented with a sentence in Chapter 5, Experiment 2. Such results suggest that WMC can explain many of the findings of reduced performance in dual-task situations. Working memory capacity also accounted for the difference between detection of syntactic and semantic errors in sentences in Chapter 5, Experiment 2. This finding suggests that semantic errors were detected less well than syntactic errors because of limited resources available for the dual-task. This interpretation was supported in Chapter 5, Experiment 3, which revealed no difference in detection of semantic and syntactic errors in the spoken sentences from Chapter 5, Experiment 2 in a single-task paradigm. This pattern of results suggests that working memory is fundamental to the concurrent processing of music and language, and is strongly tied to syntactic processing in a dual-task situation.

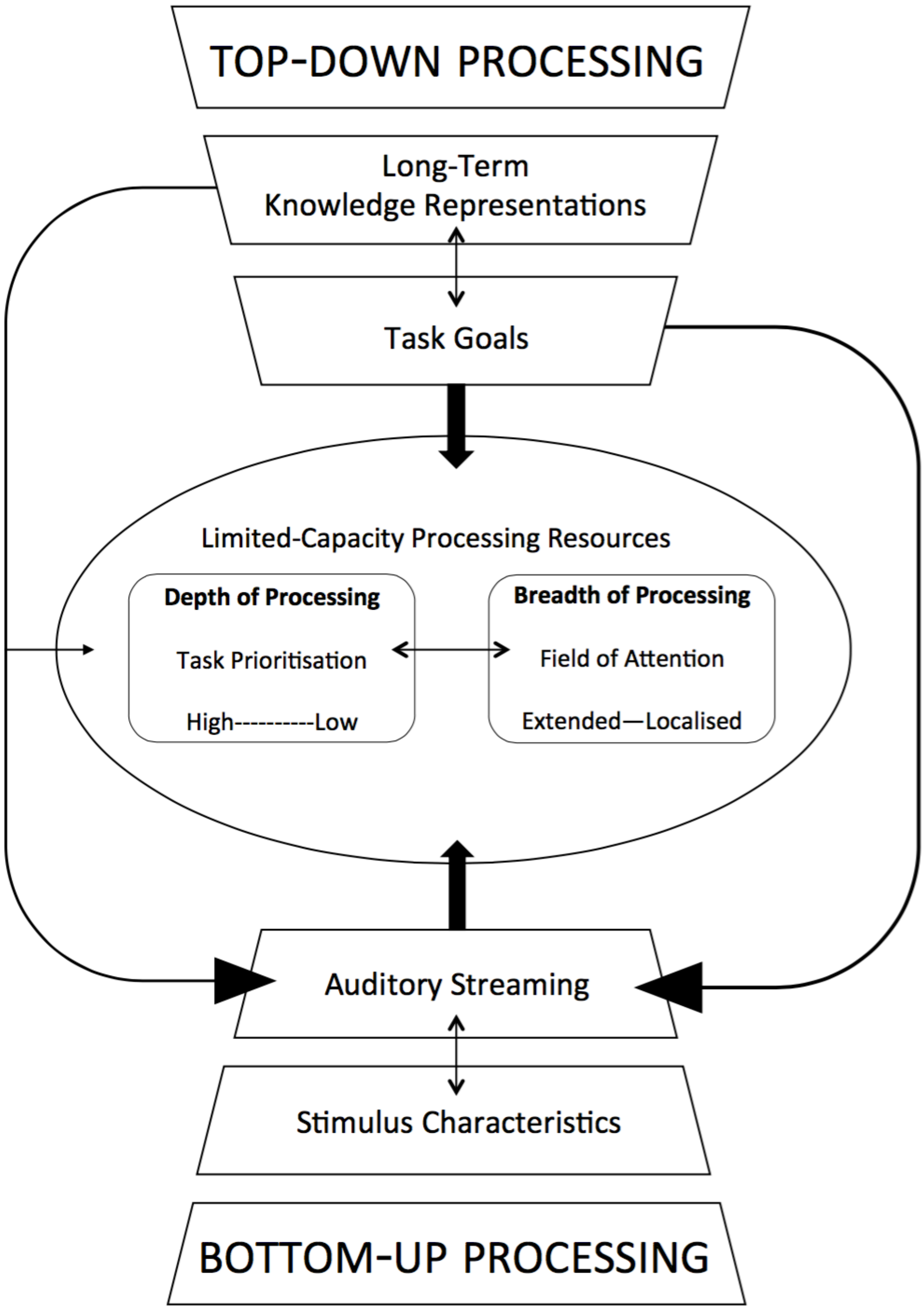

Figure 1. Visual representation of the competitive attention and prioritisation (CAP) model. Top-down and bottom-up processes influence the allocation of limited-capacity processing resources based on task prioritisation and field of attention. Task prioritisation in a dual-stimulus situation is on a continuum (low to high prioritisation). Field of attention can either be extended or localised, and is flexible depending on task demands.

Depth of processing (task prioritisation). To successfully complete two tasks at the same time, it is likely that participants cognitively prioritise one of the tasks. This process is referred to as task prioritisation. Processing resources may be allocated to each task and compete for resources depending on current task goals and the level of attention engaged by the stimulus. I propose a continuum between high and low prioritisation, often resulting in a *primary* and a *secondary* task. Resources could also be shared amongst three or more tasks. However, for the current discussion and in relation to the current results, I will focus the discussion on dual-stimulus situations. Primary tasks could be explicitly specified by the experimenter (e.g., focus on one task—such as sentence recall in Chapter 2, Experiment 1), or inferred implicitly (e.g., participant is given feedback on only one task—such as the same-different task in Chapter 5, Experiment 1; or one task is more difficult than the other—such as language comprehension in Chapter 4, Experiments 1 and 2). Primary tasks are highly prioritised, and allocated the majority of attentional resources. It should be noted that the experimenter-defined primary task might not be the same as the task that is prioritised cognitively (e.g., if the intended secondary task is particularly challenging).

Prioritisation could also occur in real-life dual-task situations. For example, an explicit task goal may be to read a book (primary task) and so any concurrent music (or background speech) would be a secondary task. The background auditory stimulus is still processed, but the majority of attention is allocated to the primary task. If the secondary task required more processing (e.g., your name occurred in the background speech, or there was an unexpected event in a background musical sequence), then the balance of task prioritisation might shift—the secondary task would be allocated more resources, and the primary task would be allocated fewer resources. Both tasks therefore compete for a limited amount of resources. There are also situations where both streams of information are equally important and are hence both prioritised. In these cases, both tasks will equally share processing resources, and performance is likely to decrease on both tasks. The continuum of high and low task prioritisation therefore reflects a dynamic and fluid allocation of resources, based on incoming stimuli and task goals.

Previous research has also distinguished between primary and secondary tasks in a dual-task situation. Kahneman (1973) suggested that participants are able to flexibly allocate attention to meet task demands, and that the more resources the primary task uses, the fewer resources are available for the secondary task, as in the current model. The distinction between primary and secondary tasks also fits into the attentional taxonomy presented by Chun, Golomb, and Turk-Browne (2011), where external attentional resources select the most relevant information for further processing. Assuming the participant is motivated to perform well, they will try to limit task interference by prioritising and allocating the most resources to the primary task. Therefore, it is possible that in demanding experimental situations, participants will not show interference on the primary task (unless there is a very sensitive measure—such as sentence recall in Chapter 2, Experiment 3), but will show interference on the secondary task (as observed for complexity judgements in Chapter 4, Experiment 2; and for semantic error detection in Chapter 5, Experiment 2). In these cases, interference may be observed in the secondary task, as there may not be enough resources to manage potential interference.

The difference in resources allocated to the primary and secondary tasks may also result in different levels of processing. Because primary tasks are allocated the most resources, the primary task stimulus is likely to be processed more deeply than the secondary task stimulus. Deep levels of processing and high prioritisation allow for a detailed analysis of the structural features of the incoming stimuli. The secondary task is allocated fewer resources, so these stimuli are likely to be processed at a more shallow level of processing. Shallow processing may result in a surface-level representation of the incoming stimuli, where syntactic structure is not fully processed. Shallow processing can be likened to the concept of *good enough* language comprehension, where listeners do not always fully process the details and syntactic structure of each incoming sentence, but rather, process it to an adequate extent for the current task (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007). Shallow processing therefore uses fewer resources, as the content is not processed at a deep level—especially if it is accompanied by a demanding primary task. If both tasks are competing for primary attention, participants are more likely to engage in task-switching tactics and to show interference on both tasks, unless they choose to prioritise one task.

Breadth of processing (field of attention). Whereas task prioritisation refers to the depth of processing resources allocated to each task in a dual-task situation, field of attention refers to the breadth of attention engaged for each stimulus. Both tasks in a dual-task situation engage a field of attention that can be extended or localised, depending on top-down task goals and bottom-up stimulus characteristics. This field of attention is flexible, and can alternate between extended and localised within a sequence if necessary. Top-down task goals would result in an extended field of attention if participants were asked to make a judgement or a response based on a *whole sequence.* For language, an extended field would be engaged if participants were asked to recall a full sentence (e.g., Chapter 2, Experiments 1 and 3) or to answer a language comprehension question (e.g., Chapter 4, Experiments 1 and 2). For music, an extended field would be engaged if participants were asked to compare two sequences for overall structure (e.g., Chapter 5, Experiment 1), recall a musical sequence (e.g., Chapter 5, Experiment 2), or to judge harmonic closure (e.g., Kunert et al., 2016).

Stimulus characteristics such as long distance dependencies within a sequence would also engage an extended field of attention even if the stimulus were not task-relevant. For example, in the sentence: *The king who the knight helped sent a gift from his castle,* an extended field of attention would be necessary to maintain and integrate words into the evolving structure. Naturalistic music listening would also engage an extended field of attention to process long distance dependencies within a musical sequence. For example, a musical sequence that starts with the tonic chord, moves through other chords in the key, and then resolves with the tonic requires long distance integration of elements. In experimental situations, if a judgement were not required (e.g., the secondary stimulus is not task relevant, as in Chapter 2, Experiment 1), then an extended field of attention would be engaged, unless there were an attention-grabbing element that might then engage a localised field of attention (e.g., Fedorenko et al., 2009; Fiveash & Pammer, 2014; Hoch et al., 2011; Koelsch et al., 2005; Slevc et al., 2009). Furthermore, syntactic and semantic violations that involve a contextual integration and/or long distance integration are also likely to engage an extended field of attention, as the entire sequence is often necessary to make a judgement (e.g., the semantic errors in Chapter 5, Experiment 2). All of these examples require the participant to extend their field of attention across the whole sequence to make a judgement or process a sequence.

A localised field of attention can also be engaged based on top-down task goals and bottom-up stimulus characteristics. Top-down task goals engage a localised field of attention if participants have to focus on *parts of the sequence;* for example, if they are asked to detect a deviant tone such as an out-of-key note or gong (e.g., Chapter 2, Experiment 3) or a timbre change (e.g., Steinbeis & Koelsch, 2008). A localised field would also be engaged if there were a bottom-up attention-grabbing element, such as an increase in loudness (e.g., Fedorenko et al., 2009; Kunert, Willems, Casasanto, Patel, & Hagoort, 2015), an out-of-key note (e.g., Fiveash & Pammer, 2014; Hoch et al., 2011; Slevc et al., 2009), or an unexpected element in the sequence. Although out-of-key elements are related to the overall tonality of the piece of music (involving long distance dependencies), when a strong sense of key is built up, an out-of-key element is immediately recognised as a violation, and is likely to engage a localised field of attention. On the other hand, an ambiguous tonal context might require an extended field of attention to determine whether the note or chord fits into the context. Furthermore, local syntactic violations in language that do not require context to determine if they are a violation (e.g., *one cats*, *two cat* in Chapter 5, Experiment 2), would also engage a localised field of attention, as they can be immediately classified as errors. The evolving syntactic representation in the sentence (similar to music) allows listeners to determine the error at a local level, requiring only a localised field of attention.

The empirical basis for CAP. The CAP model suggests that interference between music and language relies on a complex interplay between task prioritisation (depth of processing) and field of attention (breadth of processing). The allocation of processing resources to different tasks (task prioritisation) and the type of attention allocated to tasks (field of attention) are based on top-down task goals and bottom-up stimulus characteristics. The sensitivity of the dependent variable is also an important factor as to whether or not syntactic interference can be observed between two syntactic sequences. These factors have emerged as important in the experiments presented throughout the thesis, and in an examination of previous syntactic interference effects.

Table 1 in Appendix C outlines the dual-stimulus experiments in the current thesis. These experiments are broken down into task prioritisation, the task itself, field of attention, level of processing, and whether or not interference was observed for each task.

As can be seen in this table, interference has been observed on the primary task, the secondary task, and on both tasks simultaneously. Chapter 2, Experiment 1 showed syntactic interference in the primary task of sentence recall when both the primary and secondary (auditory processing) tasks involved an extended field of attention. In addition, Chapter 2, Experiment 3 revealed syntactic interference in the primary task (sentence recall, extended field of attention), when the secondary task (target detection) engaged a localised field of attention (detect an out-of-key note or a gong sound). The secondary task in this experiment also produced syntactic interference, as participants were significantly worse at detecting out-of-key notes when concurrently processing sentences compared to word-lists, an effect that did not occur for gong detection. Chapter 2, Experiment 3 therefore revealed a bidirectional syntactic interference effect in both the primary and secondary tasks. I suggest that interference was found in both tasks because the dependent variable of language recall was very sensitive, and the secondary task of target detection required a deep level of processing. Therefore, I predict that interference should be observed in the primary task when both tasks require an extended field of attention and the dependent variable is sensitive (e.g., Chapter 2, Experiment 1). Interference should also be observed in both the primary and secondary tasks when the primary task has a sensitive dependent variable, and both tasks require a deep level of processing (e.g., Chapter 2, Experiment 3).

Chapter 4 included two experiments where language comprehension was measured while participants were listening to basic and complex music and environmental stimuli. Interference was not observed on the primary task (language comprehension, extended field of attention) in either Experiment 1 or Experiment 2. The secondary task in Experiment 1 was to judge whether auditory stimuli were musical or environmental. It is difficult to determine if participants engaged an extended or localised field of attention for this simple judgement, as it could be performed with an extended field of attention (e.g., judge whole sequence to determine stimuli type), or with a localised field of attention (e.g., focus on one element to determine stimuli type). In either case, interference was not observed in Experiment 1 on the secondary task, likely due to ceiling effects. The secondary task in Experiment 2 required participants to judge the complexity of auditory stimuli. This judgement is likely to have required an extended field of attention to judge complexity of the whole sequence, and syntactic interference was observed on this task. Musical complexity judgements were reduced when participants were performing a language comprehension task compared to when they were performing a visuospatial search task. This result suggests that interference can be observed in the secondary task when both tasks require an extended field of attention, and the dependent variable in the primary task is robust to interference.

In Chapter 5, I presented two experiments where the primary tasks were music tasks that required an extended field of attention (melodic same-different task, melody recall task). The secondary task for both experiments was the detection of errors (either syntactic or semantic) in the sentence stimuli. It appears that the different types of errors in the stimuli from Chapter 5, Experiments 1 and 2 engaged different fields of attention. It is likely that the syntactic errors engaged a localised field of attention, as the errors did not require long distance integration. The semantic error and no error sentences on the other hand may have engaged an extended field of attention, as the whole sentence had to be analysed to make a judgement. The secondary task therefore involved a changing field of attention, as participants were unaware of the type of error that would occur when they began to process the sentence. Chapter 5, Experiment 1 did not show interference on the primary task; however, correlations with WMC suggested that syntactic errors resulted in different processing than semantic errors and no errors. Although detection of semantic errors in the secondary task was lower than detection of syntactic errors or correct identification of no errors, a subsequent experiment (Chapter 5, Experiment 3) suggested that this difference was inherent to the stimuli.

Chapter 5, Experiment 2 was designed to clarify the findings of Chapter 5, Experiment 1. The primary task did not reveal a difference for total melody recall depending on sentence error type. A closer examination of recall for the last two notes of the melodies (when the language errors were presented) showed a recency effect in serial recall (enhanced recall for the final note) for the syntactic error and melody only conditions, as might be expected based on serial recall studies (Roberts, 1986). However, there was a *decrease* in performance from the penultimate to final position for the semantic error and no error conditions. This difference suggested that interference occurred in the primary task at the point of the error for the semantic error and no error conditions (requiring an extended field of attention) but not for the syntactic error condition (requiring a localised field of attention). These results suggest that overlapping fields of attention were required to show interference on a relatively simple music task. In addition, Chapter 5, Experiment 2 showed that participants were significantly worse at detecting semantic errors in spoken sentences compared to syntactic errors or no error in the secondary task. Chapter 5, Experiment 3 showed that this difference was driven by the dual-task nature of the experiment, as there was no difference in error detection between semantic and syntactic errors in a single-task paradigm with the same stimuli.

The pattern of results in Chapter 5, Experiment 2 suggests that (a) interference in the primary, melody recall task was observed at the point of error for the semantic error and no error conditions compared to the syntactic error and melody only conditions, (b) interference in recall for the last two notes may result from sentence reanalysis that only occurred when processing sentences with no error or a semantic error (an extended field of attention), and (c) interference was observed in the secondary task, with decreased *semantic* error detection for sentences. These results suggest that interference may not be at the level of syntax specifically, but may be observed when two tasks require an extended field of attention. It is possible that overlap in processing between music and language may be related to processes of reanalysis and integration (engaging a wide field of attention), as suggested by Slevc and Okada (2014) and Van de Cavey et al. (2017), rather than syntactic processing specifically.

The CAP model also predicts that interference will be observed when two localised fields of attention are engaged simultaneously. Studies by Koelsch et al. (2005), Steinbeis and Koelsch (2008), and Hoch et al. (2011) provide evidence for this suggestion. These studies all paired an out-of-key or unexpected chord with a syntactic violation (gender disagreement) or a semantic anomaly (less expected word). Interference was found with the syntactic violations but not the semantic violations. The local gender violations in these instances are unambiguous syntactic errors, and appear less conducive to reanalysis than ambiguous semantic anomalies. An example from Hoch et al. (2011) is the sentence, translated from French: *The nasty dog is sleeping in the kennel*. French words have particular genders assigned to them. In the syntactic violation condition, the word *kennel* takes on an unexpected gender, resulting in an unambiguous syntactic violation. A semantic anomaly on the other hand involves the replacement of the word *kennel* with the word *tent*. *The nasty dog is sleeping in the tent* is entirely plausible, and is unlikely to result in a localised field of attention when encountering the word *tent*. Therefore, it is highly possible that previous studies observed interference between syntactic violations in language and out-of-key chords in music because both tasks required a localised field of attention at exactly the same time. On the other hand, the semantic violation may have engaged a wide field of attention that did not interfere with the out-of-key chord. This explanation could explain why previous research has observed interference for syntactic errors but not semantic errors—not because of specific syntactic overlap, but rather, because syntactic violations are more salient than semantic anomalies and engage a different field of attention.

Out-of-key notes also appear to engage a localised field of attention and to draw attention away from the processing of complex, garden path sentences. Both syntactic and semantic garden path sentences are read more slowly when the disambiguating word is presented at the same time as an out-of-key chord (Perruchet & Poulin-Charronnat, 2013; Slevc et al., 2009). In the study by Slevc et al. (2009), reading time of syntactic garden path sentences was not affected when a chord was played with a different timbre, or when sentences with a semantic anomaly were paired with an out-of-key chord. However, as discussed throughout this thesis, it is likely that out-of-key chords are more salient than chords played with a different timbre (Tillmann & Bigand, 2015), and semantic anomalies are unlikely to engage the same level of processing necessary to comprehend a garden path sentence. It is also important to note that neither Slevc et al. (2009) nor Perruchet and Poulin-Charronnat (2013) observed interference on the primary task of language comprehension—interference was only observed for reading times of sections of sentences. As such, these studies are in line with the current findings since interference was not observed on the primary task when language comprehension was the dependent variable. Further, in the experiments reported by Slevc et al. (2009) and Perruchet and Poulin-Charronnat (2013), participants only had to perform one task. As there was no secondary task, it is likely that the majority of resources were allocated to language comprehension, which was therefore robust to interference. However, this interpretation does not support the findings of Fedorenko et al. (2009), who did find interference with an out-of-key note on language comprehension of a sung sequence. The combined language and melody in one melodic stream, and the sped up nature of the auditory stimuli may have resulted in this difference. Therefore, interference can be observed with the simultaneous presentation of stimuli engaging a localised (e.g., out-of-key chord) and an extended (e.g., garden path sentence) field of attention, when the stimulus engaging the extended field of attention requires deep processing and reanalysis.

Both Kunert et al. (2016) and Van de Cavey et al. (2017) measured performance on a primary music task (i.e., harmonic closure judgements and probe-tone judgements), and had participants respond to questions on a secondary language task at points throughout the experiment. Secondary task judgements were not analysed depending on condition, but participants were paying attention to both tasks. Both papers reported interference in the primary task when the music and language stimuli involved an extended field of attention. Kunert et al. (2016) observed reduced harmonic closure judgements for chord sequences that were paired with garden path sentences compared to sentences that were unambiguous. Although both types of sentence required an extended field of attention, the garden path sentences would have required reanalysis and a deeper level of processing than an unambiguous sentence, thereby resulting in higher task prioritisation of these more complex sentences, and less resources available for the primary task of harmonic closure judgements. Van de Cavey et al. (2017) observed that phrase boundaries in pitch sequences were processed to a lesser extent when participants were concurrently reading garden path sentences compared to sentences with a syntactic error. Syntactic errors, as discussed, are likely to engage a localised field of attention, and therefore may not interfere as strongly with the extended field of attention required for the music task. Both of these results can therefore be explained by the task prioritisation and field of attention aspects of the CAP model. Kunert et al. (2016) and Van de Cavey et al. (2017) suggested that their results occurred because music and language draw on overlapping syntactic integration processes. This suggestion is compatible with the CAP model, and indeed, may be the source of overlap with two concurrent tasks requiring an extended field of attention. However, the CAP model can also explain interference between two local violations (e.g., a syntactic error in a sentence and an out-of-key chord) based on overlap in localised fields of attention.

The CAP model accounts for the pattern of results presented in the current thesis, and can be applied successfully to instances of syntactic interference and non-interference in the existing music and language syntactic processing literature. The experiments presented in this thesis are critical to uncovering the nature of shared processing between music and language. By investigating performance on music and language tasks simultaneously, I was able to uncover a complex interplay between task prioritisation and field of attention which is compatible with previous findings in the literature. Overlapping fields of attention can explain instances of interference at a local level (involving violations of syntactic structure), and overlap at an extended level (involving processes of reanalysis or processing without violations). However, interference is not only at the level of field of attention. Task prioritisation with an attached continuum of high and low prioritisation helps to explain instances where interference is observed without overlapping fields of attention. In these instances, the depth of processing required for each task, and the sensitivity of the dependent variable are important elements that influence whether interference is observed. I suggest that task goals and stimulus characteristics drive a complex interplay between field of attention (extended, localised) and task prioritisation (high-low) that is based on top-down task goals and bottom-up stimulus characteristics. Further, the CAP model underlines the importance of measuring processing of both stimuli within a dual-stimulus paradigm to elucidate these connections.

The CAP Model: Challenges and Prospects

The CAP model was designed to account for patterns of syntactic interference and non-interference between music and language, and to place syntactic processing within a larger cognitive model. The CAP model incorporates a number of important influences on the concurrent processing of two syntactic streams of information, and makes a number of testable predictions that can influence the design of future research. Specifically, the CAP model predicts that interference between music and language should be greatest when (a) both music and language processing require a similar field of attention (e.g., both localised, both extended), (b) both tasks are highly prioritised and involve a deep level of processing, and (c) participants are required to make a response during or immediately after potential interference. The CAP model also predicts that interference will not always be observed on the primary task, especially when performance on the primary task is measured with a dependent variable that is not particularly sensitive, such as language comprehension. In these situations, interference is predicted on the secondary task, as fewer resources are available to process a non-prioritised channel of information. Below I outline how future studies might design appropriate stimuli and tasks in order to test the validity and reliability of such predictions.

Considerations for stimulus design. The first prediction of the CAP model is that interference should occur when concurrent music and language stimuli require a similar field of attention. To test this prediction, stimuli should be designed in a manner that engages either a localised field of attention for both streams, or an extended field of attention for both streams. Instances of task demands that require an extended field of attention have been discussed in this thesis (for music: same-different judgements, melody recall, harmonic closure judgements; for language: sentence recall, language comprehension).

When a syntactic violation is introduced, a localised field of attention should be elicited. To engage a localised field of attention without introducing a syntactic violation, participants could be asked to focus on *specific aspects* of the music or language sequence, rather than the sequence as a whole. An example in language might be to present a written sentence on the screen, and ask participants to detect subtle changes in font or text size. Another example would be to present an auditory sentence, and ask participants to detect slight differences in voice timbre, or whether a certain word occurs in the sentence. An example in music might be to ask participants to detect targets in the auditory stimulus (as in Chapter 2, Experiment 3), or to detect subtle changes in instrument timbre or loudness. When participants are concurrently engaging a localised field of attention in music and language, interference should be observed.

The second main prediction is that interference should be observed when both tasks are highly prioritised and hence involve a deep level of processing. One of the recurrent findings in this set of experiments was that unusual stimuli resulted in a processing cost beyond what was predicted based on the manipulation. This cost likely occurred because the unusual stimuli required a deeper level of processing and may have been more highly prioritised than other stimuli. For example, the scrambled stimuli in Chapter 2, Experiment 1 did not lead to a non-syntactic sequence of notes as predicted, and may have been prioritised. Considering it was likely that the scrambled melodies were perceived as one auditory stream (proximal pitches, proximal note distances, and the same timbre), it is possible that participants prioritised scrambled sequences in an attempt to perceive and encode structure, hence recruiting even more syntactic processing resources. Likewise, in a melody recall pilot study presented in Appendix A, participants were presented with reversed spoken sentences that were designed to eliminate syntactic structure. Instead, the unusual nature of these stimuli resulted in poorer performance than expected. The environmental sound chunks (Chapter 1, Experiment 3; Chapter 4) were also unusual for listeners, and many participants spontaneously described the stimuli as strange. Thus, it is challenging to create auditory sequences that do not contain syntactic structure because any large deviations from typical stimuli may result in deeper processing, and hence interference. An alternative approach is to disrupt the auditory stream within sequences, which should make the syntactic representation less coherent.

The three-timbre stimuli presented in Chapters 2 and 3 included alternating piano, acoustic guitar, and vibraphone musical instrument digital interface (MIDI) instruments within a melody, and three different voices within a sentence. I have suggested that by alternating these timbres, auditory streaming processes were disrupted, and syntactic structure processing was reduced. Based on this suggestion, any stimulus that disrupts auditory streaming should result in similar effects. Future research could investigate whether this effect occurs with other timbres and different sound locations (e.g., alternating the direction each note or chord is coming from), as these manipulations should also interrupt auditory streaming. It would also be valuable to investigate the three-timbre effect in spoken sentences to a greater extent, by using more distinct voices, and more trials per condition for the auditory sentences. Interrupting the auditory stream in a syntactic sequence is therefore one direction forward in designing stimuli that can investigate shared processing between music and language without violations of syntactic structure, and without engaging deep syntactic processing.

Considerations for task design. The experiments presented in this thesis show that the choice of experimental tasks and related dependent variables is crucial to whether or not interference is observed between music and language, as they influence task prioritisation and field of attention. Sentence recall appears to be a sensitive variable in measuring processing costs (e.g., in Chapter 2, Experiments 1 and 3). In contrast, language comprehension appears relatively robust to syntactic interference (e.g., in Chapter 4). This difference between language comprehension and language recall is likely to occur because participants are able to understand the "gist" of the sentence and process it for meaning, even when concurrently processing other stimuli. Recalling a whole sentence on the other hand requires a deep level of processing to maintain the sentence in working memory before repeating it out loud. To elicit interference in language comprehension with concurrent auditory processing might therefore require a demanding secondary task, or the language comprehension task to be made more demanding (e.g., presentation sped up as in Fedorenko et al., 2009).

In Chapter 5, I investigated whether syntactic interference could be observed from language to music. In designing these experiments, it became apparent that the music task had to be quite simple to be performed by non-musicians. Future research could potentially increase the sensitivity of this design by asking musicians to perform a complex music task. A complex music task should engage syntactic processing to a greater extent, and is therefore more likely to result in interference. My aim was to understand syntactic processing among both musicians and non-musicians, but it is possible that the music task required to achieve this aim was too simple to produce syntactic interference. Future research could therefore investigate language to music interference in musicians by asking participants to compare two structurally complex pieces of music, or to recall a more complex piece of music. Further, musicians could be asked to reproduce a musical sequence on their instrument, as such a design would align more closely with language recall experiments. Complex music tasks may result in deeper syntactic processing that is required to show interference.

Future directions for the CAP model. Future research should continue to test the CAP model based on the predictions outlined above. The relative contributions of and connections between task prioritisation and field of attention in relation to limited-capacity processing resources should also be further investigated. An important step is to quantify levels of task prioritisation across two concurrent stimulus streams. One way to measure task prioritisation is to implement a within-subjects design and directly compare performance on each task individually compared to performance on both tasks together. This comparison would allow for a direct test of the effect of a dual-task situation on the primary and secondary tasks. By carefully manipulating different aspects of the CAP model such as depth of processing (task prioritisation) and breadth of processing (field of attention), the relative contributions of each aspect of shared processing resources can be elucidated.

Future research should also adopt a more systematic and rigorous approach to investigating the role of working memory and attention in the concurrent processing of music and language. The findings presented in this thesis in relation to WMC were correlational, and hence provide only preliminary evidence for the role of WMC in syntactic processing. To directly measure the contribution of working memory to syntactic processing, future investigations could measure performance on concurrent music and language tasks while participants are under either a low working memory load or a high working memory load (e.g., as in Dalton, Santangelo, & Spence, 2009). By directly manipulating working memory load, the contribution of working memory to syntactic processing in music and language could be measured. Direct manipulations of attention as in Maidhof and Koelsch (2011) and Loui, Grent-'t-Jong, Torpey, and Woldorff (2005) could also be used to influence task prioritisation and to ensure particular fields of attention when processing stimuli. Testing predictions of the CAP model can therefore further increase our understanding of the concurrent processing of music and language.

Concluding Remarks

The study of music and language—two of the most complex syntactic systems comprehended and produced by humans—can provide an insight into how the brain processes meaningful syntactic information. Furthermore, the study of how the brain processes music and language at the same time can provide an insight into the limits and nature of our cognitive processing resources. The current thesis aimed to test theories of shared syntactic processing between music and language. Throughout the course of experimentation, it became clear that the current models of shared syntactic processing could not account for a number of dual-stimulus situations, and a new model was needed to understand how the brain processes music and language concurrently when there are no violations of syntactic structure. The CAP model was proposed, which incorporates bottom-up processes of stimulus characteristics and auditory streaming as well as top-down processes of long-term knowledge representations and task goals. These processes influence the allocation of limited-capacity processing resources available for the concurrent processing of music and language based on task prioritisation (depth of processing) and field of attention (breadth of processing). This thesis extends the understanding of shared syntactic processing between music and language, and provides a model of shared processing that takes into account a number of potential influences that affect the concurrent processing of two syntactic streams of information. The CAP model and the experiments presented within this thesis can inform future research into this intriguing area.

References

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex, 21*(7), 1507-1518. doi:10.1093/cercor/bhq198

Baddeley, A., & Hitch, G. (1974). Working memory. In H. B. Gordon (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Braze, D., Shankweiler, D., Ni, W., & Palumbo, L. C. (2002). Readers' eye movements distinguish anomalies of form and content. *Journal of Psycholinguistic Research, 31*(1), 25-44. doi:10.1023/a:1014324220455

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology, 62*, 73-101. doi:10.1146/annurev.psych.093008.100427

Dalton, P., Santangelo, V., & Spence, C. (2009). The role of working memory in auditory selective attention. *The Quarterly Journal of Experimental Psychology, 62*(11), 2126-2132. doi: 10.1080/17470210903023646

Eerola, T. (2016). Expectancy-violation and information-theoretic models of melodic complexity. *Empirical Musicology Review, 11*(1), 1-17. doi:10.18061/emr.v11i1.4836

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences, 108*(39), 16428-16433. doi:10.1073/pnas.1112937108

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition, 37*(1), 1-9. doi:10.3758/MC.37.1.1

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in langauge comprehension. *Current Directions in Psychological Science, 11*(1), 11-15.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass, 1*(1-2), 71-83. doi:10.1111/j.1749-818X.2007.00007.x

Fiebach, C., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during pasing: The processing of German WH-questions. *Journal of Memory and Language, 47*, 250-272.

Fiveash, A., & Pammer, K. (2014). Music and language: Do they draw on similar syntactic working memory resources? *Psychology of Music, 42*(2), 190-209. doi:10.1177/0305735612463949

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84. doi:10.1016/S1364-6613(00)01839-8

Hoch, L., Poulin-Charronnat, B., & Tillmann, B. (2011). The influence of task-irrelevant music on language processing: Syntactic and semantic structures. *Frontiers in Psychology, 2*, 112. doi:10.3389/fpsyg.2011.00112

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(6), 1050-1061.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2010). Auditory distraction and serial memory: The avoidable and the ineluctable. *Noise Health, 12*(49), 201-209. doi:10.4103/1463-1741.70497

Kahneman, D. (1973). *Attention and effort*. New Jersey, USA: Prentice-Hall Inc.

Koelsch, S. (2013). *Brain and music*. Oxford, UK: John Wiley & Sons.

Koelsch, S., Gunter, T., Schroger, E., Tervaniemi, M., Sammler, D., & Friederici, A. D. (2001). Differentiating ERAN and MMN: An ERP study. *NeuroReport, 12*(7), 1385-1389.

Koelsch, S., Gunter, T., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and music: An ERP study. *Journal of Cognitive Neuroscience, 17*(10), 1565-1577.

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS One, 10*(11), e0141069. doi:10.1371/journal.pone.0141069

Kunert, R., Willems, R. M., & Hagoort, P. (2016). Language influences music harmony perception: Effects of shared syntactic integration resources beyond attention. *Royal Society Open Science, 3*(2). doi:10.1098/rsos.150685

Levitin, D. J., & Menon, V. (2003). Musical structure procesed in "language" areas of the brain: A possible role for Brodmann area 47 in temporal coherence. *NeuroImage, 20*, 2142-2152. doi:10.1016/S1053-8119(03)00482-8

Loui, P., Grent-'t-Jong, T., Torpey, D., & Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Cognitive Brain Research, 25*(3), 678-687. doi:10.1016/j.cogbrainres.2005.08.019

Maidhof, C., & Koelsch, S. (2011). Effects of selective attention on syntax processing in music and language. *Journal of Cognitive Neuroscience, 23*(9), 2252-2267.

Pashler, H., & Johnston, J. C. (1998). Attentional limirations in dual-task performance. In H. Pashler (Ed.), *Attention*. Hove, United Kingdom: Psychology Press.

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

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733.

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychonomic Bulletin & Review, 20*(2), 310-317. doi:10.3758/s13423-012-0344-5

Roberts, L. A. (1986). Modality and suffix effects in memory for melodic and harmonic musical materials. *Cognitive Psychology, 18*(2), 123-157. doi:10.1016/0010-0285(86)90010-1

Slevc, L. R., & Okada, B. M. (2014). Processing structure in language and music: A case for shared reliance on cognitive control. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-014-0712-4

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review, 16*(2), 374-381. doi:10.3758/16.2.374

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18*(5), 1169-1178. doi:10.1093/cercor/bhm149

Tillmann, B., & Bigand, E. (2001). Global context effect in normal and scrambled musical sequences. *Journal of Experimental Psychology: Human Perception and Performance, 27*(5), 1185-1196.

Tillmann, B., & Bigand, E. (2015). A commentary on "A commentary on: 'Neural overlap in processing music and speech'". *Frontiers in Human Neuroscience, 9*. doi:10.3389/fnhum.2015.00491

Van de Cavey, J., Severens, E., & Hartsuiker, R. J. (2017). Shared structuring resources across domains: Double task effects from linguistic processing on the structural integration of pitch sequences. *The Quarterly Journal of Experimental Psychology, 70*(8), 1633-1645. doi:10.1080/17470218.2016.1195852

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177. doi:10.1080/14639220210123806

Appendices

Appendix A: Melody Recall Pilot Study

Appendix A

Melody Recall Pilot Study

The melody recall pilot study was designed to investigate whether syntax in spoken sentences interferes with the recall of melodies. Spoken complex sentences (with syntax) and reversed versions of these sentences (without syntax) were presented to participants at the same time as a simple melody. Based on research by Williamson, Baddeley, and Hitch (2010) described in Chapter 5, the current pilot study used seven-note melodies with three distinct notes that were seven semitones (a perfect fifth) apart. After testing 12 participants, it became clear that the seven-note melodies were producing floor effects. Thus, melodies were shortened to six notes for four participants. Recall was increased for the six-note melodies, and the pattern of results was the same, so the data for these two subsets of participants were merged for the analysis. Working memory capacity (WMC) was also measured, as we expected WMC to be positively correlated with melody recall. We also predicted that WMC would account for any syntactic interference effects present because of the integral role of working memory to syntactic processing. It was predicted that the complex sentences would result in poorer melody recall than reversed sentences because complex sentences would induce syntactic interference.

Method

Participants. Sixteen undergraduate psychology students from Macquarie University participated in this study for course credit. Twelve participants (seven females and five males, 10 right handed, two left handed; M_{age} = 20.75, *SD* = 5.89) recalled the seven-note stimuli. For these participants, 10 indicated some level of musical training, with an average of 3.3 years of private lessons and 7.2 years of combined private and informal experience. Three participants indicated they were currently musically active, and four considered themselves either a musician, or “somewhat” a musician. Four participants (all female, all right handed, $M_{\text{age}} = 21.75$ years, $SD = 3.59$) recalled the six-note stimuli. Three of these participants indicated some musical training, with an average of 5.8 years of private lessons, and 9.9 years of combined private and informal experience. Two of these participants indicated they were currently musically active, and three considered themselves as “somewhat” musicians.

All 16 participants spoke English as their first language and were born in either Australia ($n = 15$) or New Zealand ($n = 1$). All participants indicated that both of their parents spoke English as their first language, and that English was spoken 100% of the time at home. None indicated any hearing or language problems, and all indicated listening to music, with an average listening time of 129 minutes per day. None reported having perfect pitch.

Design. In the pilot study, participants listened to and then recalled short melodies that were presented (a) on their own (baseline), (b) with complex sentences (syntax), or (c) with reversed sentences (no syntax). Scenarios were created to provide context for these sentences and make them more meaningful to encourage close listening. Scenarios were presented before each group of stimuli, and scenario presentation was randomised for each participant. Sentences, reversed sentences, and melody-only trials were randomised within each scenario for each participant. The experiment was programmed and presented with Matlab (version 2016b; Mathworks) and Psychtoolbox (Brainard, 1997).

Appendix A: Melody Recall Pilot Study

Language stimuli. Language stimuli were recorded in a sound proof booth with an AKG 535 condenser vocal microphone and recorded into Audacity. All stimuli were spoken by an Australian female who was a native-English speaker. To ensure sentences were meaningful to participants, they were grouped into themes, and prefaced by a short scenario that introduced the scene and the characters in the sentences. Scenarios consisted of 3-4 sentences and were designed to provide a context for the following sentences. There were six scenarios in total, with the themes of: fantasy, crime, wedding, school, zoo, and kitchen. The fantasy scenario was: *Once upon a time, there was a King who wanted to throw a party in his castle. He invited many people, including his family—the Queen, the Prince, his sister, and niece. He also invited other guests including a knight from a neighbouring land. There was music, good food, and wine.* Scenarios were read with natural prosody in a story telling voice by the same speaker who read the sentences. Within each scenario, there were five sentences and five reversed sentences, resulting in thirty sentences and thirty reversed sentences in total.

Twelve-word object-extracted (complex) sentences were adapted from Fedorenko et al. (2009) and recorded with natural prosody. An example of a sentence is: *The guest who the queen kissed brought a cake to the party*. Before recording each sentence, the speaker listened to seven notes (all the same pitch) to have an understanding of the length each sentence should be. All sentences were between 3-4 seconds long when recorded.

Reversed sentences consisted of the same object-extracted sentences as above spoken in reversed order, for example: *Party the to cake a brought kissed queen the who guest the*. The speaker practiced saying these sentences, and recorded them after listening to the length of a seven-note melody, as in the sentence condition. Because reversed sentences did not flow as naturally as forward sentences, they had to be spoken slightly faster to fill the same time frame.

After recording, the sentences and reversed sentences were imported into Audacity. Each sentence and reversed sentence was manipulated so there were no silences before or after the utterance. Tempo was then altered in Audacity using the change tempo function that changes tempo without changing pitch. All language stimuli were altered to 3.72 seconds to correspond to the seven-note melody length. Sentences were carefully altered to ensure no artificial sounds were introduced into the spoken language. If altering the tempo introduced artefacts, then the original sentence was altered slightly by either inserting more space between words or removing some silence. This procedure ensured that altered audio still sounded natural.

Melodic stimuli. To create the seven-note melodies, three notes were selected in the key of C that were easily distinguishable, a perfect 5^{th} (seven semitones) apart, and tonal. These were C_4 (the tonic note), G_4 (the fifth), and D_5 (the second, in the next octave). Ninety sequences were randomly created with these three notes using Matlab (version 2016b; Mathworks), with the restrictions that: (a) each successive note was distinct, (b) each melody contained all three notes, and (c) an equal proportion of melodies started on each note. Musical instrument digital interface (MIDI) files were created with MuseScore (version 2.0.2), imported into GarageBand, and exported with MIDI instrument Steinway grand piano at a tempo of 120bpm. Melodies were then imported into Audacity, and normalised in loudness with respect to the language stimuli. Six-note stimuli were created the same way, however the tempo was decreased to 100bpm to match sentence length.

Final stimuli. Once all sentences, reversed sentences, and melodies were the same length and normalised in respect to loudness levels, they were randomly combined. A random number generator allocated melodies to one of three conditions: melody only, melody paired with sentences, and melody paired with reversed sentences. Two different randomisations were created to ensure any differences were not pairing specific.

Procedure. Participants gave informed written consent and filled out a music education and preference questionnaire. They then received pitch training. Pitch training consisted of listening and pitch judgement exercises. First, participants heard a C chord, followed by the three notes (C, G, D) they would be discriminating. These notes were described as "low", "medium", and "high". The three notes were played 10 times in a row to familiarise participants with the distances between each note. Participants then had 10 trials where they heard a note and indicated whether it was low, medium, or high. For these trials, participants first heard the C chord, and then the target note. Feedback on this task indicated whether they were correct or incorrect. If incorrect, the correct answer was provided. If participants were not confident in this task they could repeat the process. Before each melody in the practice and experiment proper, participants were presented with a C chord, and the three notes (C, G, D) played at 120bpm. Participants had two practices recalling the melodies using the 2, 5, and 8 keys on the right keypad of the keyboard, corresponding to the low, medium, and high notes. Participants were then informed that in some trials speech would be presented with the melodies, that the sentences could either be meaningful or meaningless, and to ignore the speech and focus on melody recall. Participants practiced melody recall under both sentence conditions. Once the participant verbally indicated that they understood the experiment, the trials began. Each set of 15 trials began with a scenario.

Appendix A: Melody Recall Pilot Study

In each trial, a fixation-cross appeared on the screen for 1 second and then the auditory stimulus played for its duration (3.72 seconds). The word "recall" then appeared on the screen, and participants were required to use the keyboard keys to replicate the melody. Participants had a break before each new scenario. Following the main experiment, participants completed the letter-number sequencing subtest of the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 2008) as a measure of WMC. This procedure is described in detail in Chapter 5. The whole procedure took approximately 45 minutes.

Scoring. Percentage correct was calculated for each of the note positions (1-7 or 1-6) based on whether participants indicated the correct note in the correct position. Average accuracy across all positions was then calculated to create a total score across the whole melody. Accuracies were calculated for each condition: melody only, paired with sentences, and paired with reversed sentences. Because the pattern of results was similar for the seven-note and six-note melodies, all participant scores were included in the analysis. Musical training was not included as a covariate because the participants who listened to the six-note melodies also had a disproportionate amount of musical training, so the results may not be reflective of true effects.

Results

Melody recall. A repeated measures ANOVA was conducted in order to examine the effects of the three conditions (melody only, sentences, reversed sentences) on melody recall. This analysis revealed a main effect of condition, $F(2, 30) = 6.64$, $p = .004$, $\eta^2 = .31$. Pairwise comparisons with Holm-Bonferroni corrected p values for three comparisons revealed that this main effect was driven by the melody only condition ($M = .59$, $SD = .16$) resulting in significantly better recall than the reversed sentences condition ($M = .53$, $SD = .17$), $t(15) = 3.96$, $p' = .003$, $d = 1.02$. There were no differences between the melody only and sentence conditions ($M = .57$, $SD = .20$), $t(15)$

= 1.14, *p'* = .27, or between the sentence and reversed sentence conditions, $t(15) = 2.29$, *p*' = .07. When WMC was included as a covariate, the main effect of condition was non-significant, $F(2, 28) = 1.24$, $p = .30$. There was no interaction between condition and WMC, $F(2, 28) = .25$, $p = .78$, and no between-subjects effect of WMC, $F(1, 14) = 1.56$, $p = .23$. This finding suggests that WMC can account for the difference in melody recall between the sentence and reversed sentence conditions.

Discussion

The melody recall pilot study was a preliminary study that showed no difference in melody recall when participants were listening only to a melody compared to when participants were concurrently listening to a melody and a sentence. However, memory for melodies was significantly worse when participants were concurrently listening to reversed sentences compared to only a melody. The reversed sentences were predicted to interfere with melody recall less than the sentences, as they did not contain syntax. However, the highly unusual nature of the stimuli may have resulted in this outcome. In addition, the spoken speech rate for reversed speech sounded faster than normal sentences, as the words did not flow together naturally. For these reasons, it was determined that the reversed stimuli were not an appropriate control condition. As mentioned previously, this pilot study also showed that performance on the seven-note melody recall task was low, and participants verbally commented on how difficult the task was. The final four participants were therefore tested on six-note melodies, and six-note melodies were subsequently used in Experiment 2 of Chapter 5. The low performance across all conditions with the seven-note stimuli could explain why there was no difference between the melody only and sentence conditions, as performance was already poor. However, it is clear that the reversed sentences were particularly disruptive to melody recall. The findings from this pilot study therefore informed the stimuli used in Experiment 2 of Chapter 5.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system. *Memory & Cognition, 37*(1), 1-9. doi:10.3758/MC.37.1.1

Mathworks (R2016b). Matlab. Massachusetts, United States: Mathworks.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale - Fourth Edition*. San Antonio, TX: Pearson.

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. doi:10.3758/MC.38.2.163

Appendix B: Musical Training Analysis, Chapter 5

Appendix B

Musical Training Analysis, Chapter 5

Appendix B presents the musical training analysis from Chapter 5, Experiments 1 and 2. It should be noted that musical training and working memory capacity (WMC) were not correlated in Experiment 1 (r = .12, p = .35) or Experiment 2 (r = -.03, p = .83), so the variance explained by musical training is separate to the variance explained by WMC presented in Chapter 5. Musical training was assessed by self-reported years of private music lessons.

Experiment 1: Musical Same-Different Task

Melodic same-different judgements. When musical training was added into the repeated measures ANOVA on the d' scores for the melodic same-different task across the three language conditions (no error, semantic error, syntactic error), there was still no main effect of language condition, $F(2, 124) = 1.04$, $p = .36$, and no interaction between language condition and musical training, $F(2, 124) = 1.08$, $p = .34$. However, there was a significant between-subjects effect of musical training, $F(1, 62) = 4.77$, $p = .03$, $\eta^2 = .07$. As musical training did not affect same-different judgements differently depending on language condition, a bivariate correlation was conducted between musical training and performance on the same-different task across all conditions. This analysis revealed that musical training was positively correlated with performance on the same-different task, (r = .27, p = .03), suggesting that participants with more musical training performed better on the musical same-different task, as would be expected for a music task.

Appendix B: Musical Training Analysis, Chapter 5

Reaction time (RT) for same-different judgements. When musical training was added as a covariate into the ANOVA investigating RT on the same-different task depending on language condition, there was still no main effect of language condition, $F(1.64, 101.40) = 1.34$, $p = .27$, and no interaction between language condition and musical training, $F(1.64, 101.40) = 1.95$, $p = .15$. Unlike the d' scores, there was no between-subjects effect of musical training on RT, $F(1, 62) = .35$, $p = .56$. The Greenhouse-Geisser correction was reported as the assumption of sphericity was violated, $\chi^2(2) = 15.38$, $p < .001$.

Sentence error detection. To investigate whether musical training affected sentence error detection (no error, semantic error, syntactic error), a repeated measures ANOVA was conducted with the added covariate of musical training. The assumption of sphericity was violated, $\chi^2(2) = 10.42$, $p = .005$, so the Greenhouse-Geisser correction was applied. There was still a main effect of language condition, $F(1.73, 107.18) = 9.91$, $p < .001$, $\eta^2 = .14$, no interaction between musical training and language condition, $F(1.73, 107.18) = 2.46$, $p = .09$, and no between-subjects effect of musical training, $F(1, 62) = .008$, $p = .93$. These results suggest that musical training did not influence error detection in language.

Sentence error detection RTs. When musical training was added as a covariate into the ANOVA measuring RTs for sentence error detection, there was still a main effect of language condition, $F(1.75, 108.77) = 4.93$, $p = .01$, $\eta^2 = .07$, no interaction between musical training and language condition, $F(1.54, 108.77) = .07$, $p = .94$, and no between-subjects effect of musical training, $F(1, 62) = .81$, $p = .37$. These results suggest that musical training does not influence reaction time for language error detection.

Appendix B: Musical Training Analysis, Chapter 5

Experiment 2: Melody Recall Task

Absolute pitch recall. When musical training was added as a covariate into the melody recall (scored with absolute pitch judgements) ANOVA, there was still a main effect of condition, $F(3, 150) = 5.86$, $p = .001$, $\eta^2 = .11$, and no interaction between condition and musical training, $F(3, 150) = 1.16$, $p = .33$. There was a between-subjects effect of musical training, $F(1, 50) = 20.6$, $p < .001$, $\eta^2 = .29$. Bivariate correlations revealed that musical training was positively correlated with melody recall in all conditions: melody only ($r = .52$, $p < .001$), no error ($r = .52$, $p < .001$), semantic error ($r = .45$, $p = .001$), and syntactic error ($r = .52$, $p < .001$), suggesting that participants with more musical training were performing better on the melody recall task compared to participants with less musical training.

Relative pitch recall. Including musical training as a covariate into the melody recall ANOVA revealed that controlling for musical training can account for much of the variance between conditions, $F(3, 150) = 2.66$, $p = .050$. There was no condition by musical training interaction, $F(3, 150) = .82$, $p = .48$; however, there was a significant between-subjects effect of musical training, $F(1, 50) = 16.20$, $p < .001$, $\eta^2 = .25$. Bivariate correlations revealed that musical training was positively correlated with melody recall across all conditions: melody only ($r = .50$, $p < .001$), no error ($r = .47$, $p < .001$), semantic error ($r = .39$, $p = .004$), and syntactic error ($r = .47$, $p < .001$), suggesting that participants with more musical training performed better on the melody recall task, but that this did not differ depending on condition.

Melody recall as a function of position. When musical training was added as a covariate into the condition by position ANOVA, the main effect of condition was marginally significant, $F(3, 150) = 2.66$, $p = .050$, and the main effect of position was still significant, $F(4, 200) = 18.69$, $p < .001$, $\eta^2 = .27$. The interaction between condition and position was still significant, $F(12, 600) = 3.24$, $p < .001$, $\eta^2 = .06$, the interaction between condition and musical training was not significant, $F(3, 150) = .82$, $p = .48$, and the interaction between position and musical training was not significant, $F(4, 200) = 1.07$, $p = .37$. There was no three-way interaction between condition, position, and musical training, $F(12, 600) = 1.23$, $p = .26$, and there was a between-subjects effect of musical training, $F(1, 50) = 16.20$, $p < .001$. These results suggest that musical training did not affect recall for melodies across positions.

Recency effect analysis. When musical training was added into the ANOVA investigating melody recall in the last two note positions depending on condition, there was no main effect of condition, $F(3, 150) = .72$, $p = .54$, no main effect of position, $F(1, 50) = 1.38$, $p = .25$, and an interaction between condition and position, $F(3, 150) = 4.99$, $p = .003$, $\eta^2 = .09$. There was no musical training by condition interaction, $F(3, 150) = 1.52$, $p = .21$, no musical training by position interaction, $F(1, 50) = 1.0$, $p = .32$, and no three-way interaction between position, condition, and musical training, $F(3, 150) = 2.4$, $p = .07$. There was a significant between-subjects effect of musical training, $F(1, 50) = 17.15$, $p < .001$, $\eta^2 = .26$.

Sentence judgements. When musical training was included into the ANOVA measuring error detection in sentences, the main effect of condition was still significant, $F(1.19, 59.57) = 26.52$, $p < .001$, $\eta^2 = .35$, and there was no interaction between condition and musical training, $F(1.19, 59.57) = 3.33$, $p = .07$. However, there was a between-subjects effect of musical training, $F(1, 50) = 5.56$, $p = .008$. Correlations between musical training and error detections reveal that musical training was positively correlated with semantic error detection ($r = .31$, $p = .03$), but not with syntactic error detection ($r = .11$, $p = .46$) or correct identification of no error ($r = .12$, $p = .40$). The Greenhouse-Geisser correction was reported as the assumption of sphericity was violated, $\chi^2(2) = 55.64$, $p < .001$. These results suggest that participants with musical training were more likely to correctly identify semantic errors.

Musical Training

The current results suggest that participants with musical training performed better on both the melodic same-different task and the melody recall task than participants without musical training, and that as musical training increased, so did performance on these tasks. Musical training was not linked to sentence error detection in Experiment 1; however, it was linked to sentence error detection in Experiment 2. This effect was largely driven by increased detection of semantic errors.

Appendix C: Dual-Stimulus Table

Appendix C

Dual-Stimulus Table

Appendix C: Dual-Stimulus Table

Appendix C: Dual-Stimulus Table

Table 1

List of Dual-Stimulus Experiments in the Thesis.

Experiment	Task Prioritisation	Task	Field of Attention	Level of Processing	Interference?
Chapter 2, Exp. 1	Primary (high)	Language recall	Extended	Deep	Yes: language recall
	Secondary (low)	Background audio (no task)	Extended	Shallow	N/A
Chapter 2, Exp. 3	Primary (high)	Language recall	Extended	Deep	Yes: language recall
	Secondary (high)	Target detection in audio	Localised	Deep	Yes: out-of-key note detection
Chapter 4, Exp. 1	Primary (high)	1) Language comprehension 2) Visuospatial search	Extended	Deep	No
	Secondary (low)	Is audio musical or environmental?	Extended?	Shallow	No: performance at ceiling
Chapter 4, Exp. 2	Primary (high)	1) Language comprehension 2) Visuospatial search	Extended	Deep	No
	Secondary (low?)	Is audio basic or complex?	Extended	Shallow?	Yes: complexity judgements
Chapter 5, Exp. 1	Primary (high)	Music: same-different	Extended	Deep	No: resources released
	Secondary (low)	Language: Error detection, last word recognition	Localised (syntactic) Extended (semantic and no error)	Shallow?	No
Chapter 5, Exp.	Primary (high)	Melody recall	Extended	Deep	Yes: any language interfered
	Secondary (low)	Language: error detection	Localised (syntactic) Extended (semantic and no error)	Shallow?	Yes: semantic errors noticed less

Appendix D: Ethics Approvals

Appendix D (ethics approvals) of this thesis has been removed as it may contain sensitive/confidential content