DOMAIN-SPECIALISED AND DOMAIN-GENERAL NEUROCOGNITIVE SUBSTRATES IN LANGUAGE AND MUSIC DOMAINS : CONTRIBUTIONS FROM ERP CORRELATES OF EXPECTANCY VIOLATIONS

Joanna Elizabeth Moodie

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2021

Identifiers to use to cite or link to this thesis:

DOI: https://doi.org/10.17630/sta/88
http://hdl.handle.net/10023/23462

Domain-specialised and domain-general neurocognitive substrates in language and music domains: Contributions from ERP correlates of expectancy violations

Joanna Elizabeth Moodie

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

February 2021

Candidate's declaration

I, Joanna Elizabeth Moodie, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 60,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2017.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date: 10/03/2021 Signature of candidate:

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 10/03/2021 Signature of supervisor:

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Joanna Elizabeth Moodie, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on all of print copy for a period of 3 years on the following ground(s):

- Publication would preclude future publication

Electronic copy

Embargo on all of electronic copy for a period of 3 years on the following ground(s):

- Publication would preclude future publication

Title and Abstract

- I require an embargo on the abstract only.

Supporting statement for embargo request

The experiments presented in this thesis are being written up for publication. Therefore, an embargo has been requested for both printed and electronic copies of this thesis.

Date: 10/03/2021 Signature of candidate:

Date: 10/03/2021 Signature of supervisor:

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Joanna Elizabeth Moodie, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date: 10/03/2021 Signature of candidate:

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force. We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below. The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

Embargo on all of electronic files for a period of 3 years on the following ground(s):

- Publication would preclude future publication

Title and Description

- I require an embargo on the title and description

Supporting statement for embargo request

The experiments presented in this thesis are being written up for publication. Therefore, an embargo has been requested for the data associated with this thesis.

Date: 10/03/2021 Signature of candidate:

Date: 10/03/2021 Signature of supervisor:

Research data access statement

Research data underpinning this thesis will be available at:

- https://doi.org/10.17630/dbcca08d–b942–4520–a3bf–db2577fb585e
- https://doi.org/10.17630/de439dee–2877–458f–b505–89969e8a067f
- https://doi.org/10.17630/de439dee–2877–458f–b505–89969e8a067f

Acknowledgements

I am hugely grateful for the support, guidance and friendship I have experienced throughout my PhD at The University of St Andrews. First, I would like to express my utmost gratitude to my supervisor, Dr Ines Jentzsch, for her invaluable guidance, thoughtful criticism and reliable mentorship. I will also be forever thankful to the Scottish Graduate School of Social Science for funding my PhD research (project reference: 1930942), and to all the people who participated in the experiments. Thank you to Dr Thomas Otto, who was my second supervisor in the second half of my PhD, for providing valuable feedback on the final draft of this thesis, to Dr Martin Campbell for his support during the first half of my PhD, and to all my previous supervisors and mentors – Dr Maggie Ellis, Professor Karen Spencer, Professor Julie Harris, Dr Stuart Ritchie and Dr Simon Cox – for showing me how fun research can be, which encouraged me to start and complete this thesis.

I would also like to take this opportunity to thank my friends and family. Thank you to Tom and my parents, for Ohana, to Lasse, for being a wonderful companion and to Fer, for being mi mejor amigo. I thank my nan, for her jokes and laughter, and my grandparents, for their delightful stories and shared love of St Andrews. Thank you to Katie, Kana, Penny, Ari, Memo, Matt, Austin, Jui and Rose for being great sources of inspiration and welcome distraction, and thank you to all the people of Deans Court 2017–2020 for the friendship, board games, debates, video games, roundnet, reeling, ceilidhs, food, quizzes, croquet and BBQs. I am so grateful to have had met and worked alongside such marvellous people, who have been endless sources of grounding joy throughout my PhD years, and from whom I have learned so much.

Table of Contents

3. Chapter 3 – ERP correlates of semantic, grammar and harmonic expectancy violations 67

Table of Figures

Table of Tables

Abstract

This thesis contributes to the interdisciplinary debate about the extent of shared and separable neurocognitive mechanisms between language and music processing. To achieve this, ERP (event-related potential) correlates of expectancy violations for major processes in each domain were investigated. The results of Chapter 3 support the distinctness of ERP effects elicited by semantic, grammar and harmonic expectancy violations (a centroparietal N400, central P600 and frontal P300, respectively). These findings support the hypothesis that these processes rely on separable neurocognitive substrates. In Chapter 4, meter violations elicit an N1 effect in the music domain, but the evidence for an N1 effect in the language domain was inconclusive. There was a P2 effect in both language and music domains, providing novel support for the hypothesis that some substrates involved in meter processing are domain-general. Finally, in Chapter 5, meter violations were presented simultaneously with semantic, grammar and harmonic violations. All aforementioned ERP effects were replicated. Crucially, there were no interactions in any ERP effects, signalling that their associated processes rely on distinct substrates. Overall, this thesis provides valuable contributions to the literature, indicating that at least some neurocognitive substrates involved in semantic, grammar and harmony processing are domain-specialised, while those involved in meter processing could be domain-general.

1. Chapter 1 – Introduction

1.1. Introduction to the debate: To what extent do language and music processing rely on shared or separable neurocognitive substrates?

For centuries, there have been complex interdisciplinary discussions debating potential links between language and music domains. The debate's origins can be dated to over two thousand years ago when Plato claimed that music's ability to "uplift the spirit" is due to its similarity to language (Neubauer, 1986). Since then, numerous theorists have hypothesised about connections between language and music. In 1781, Rosseau suggested that the first language was a type of song (Scott, 2000), and in 1848, Longfellow wrote that "music is the universal language of mankind" (Longfellow, 1848). Indeed, language and music exist in every known human society (Nettl, 2000).

The debate was advanced in 1871, when Darwin proposed that language and music share a single evolutionary precursor (Darwin, 1871), thus, suggesting a tangible theory to explain apparent links between language and music domains. Some academics have since hypothesised that roughly 200,000 years ago in hominids, a single communication-based system split into two systems: language and music (Mithen, 2007). Others suggest that music was an evolutionary precursor to language (Lynch, 1996). Although the specifics of this idea are debated, see Ross (2009) for a review, it provides foundations for hypothesising about links between language and music processing. Shared neurocognitive mechanisms could be due to their shared evolutionary past. Likewise, separation could perhaps reflect 200,000 years of independent evolution. The debate is ongoing. What Peretz et al. (2015) wrote remains true – the question of overlap of neural mechanisms between language and music domains "must still be considered as an open question" (p. 15).

There are clear distinctions between language and music. Language is primarily an everyday communication system, whereas music mainly functions as a source of entertainment or personal expression. Therefore, language is more precise in its use and meanings than music and tends to have more standardised interpretations. Moreover, there are components of language, such as nouns and verbs that have no musical counterparts. Likewise, music relies more on melodies, timbres and intricate rhythmic patterns compared to language. People tend to be more proficient in language production than music production. However, perhaps this is because language is more important for functioning in everyday life, resulting in increased training of language skills. Therefore, the tendency for differences in proficiency between language and music do not necessarily imply differences in underlying neurocognitive mechanisms. Indeed, evidence suggests that, in infancy, the speed of learning and production in language and music domains match (Brandt et al., 2012)

There are striking similarities between language and music. As mentioned above, they are both suggested to be human universals (Nettl, 2000), and all known societies put their words to music to create songs (Mehr et al., 2019). Both unfold over time and are structured hierarchically, using smaller elements to form more complex structures. In language, phonemes form words, which form sentences. Likewise, in music, notes form tones and chords which form melodies and harmonic progressions. Implicit learning in both domains appears to be hierarchical, too (Ettlinger et al., 2011). In language, infants first discriminate between vowel sounds, then phonemes and then words. They learn vocabulary passively and, gradually, language is understood. Similarly, in the music domain, infants first discriminate between timbres and notes, then melodies and then establish a sense of harmony (Cohrdes et al., 2016). Like language, music plays a crucial role in children's cognitive, emotional, and social development (Trehub, 2003).

Early theories of neural comparisons between language and music domains emphasised their differences. For example, Bever and Chiarello (1974) suggested that the brain's left hemisphere specialises in language and the right hemisphere specialises in music. Fodor (1983) proposed a modular approach, suggesting that the brain regions involved in language and music processing are entirely distinct. In hindsight, these models are over-simplistic. As discussed below, there appear to be closely related neurocognitive substrates involved in language and music domains, with intricate arrangements of sub-mechanisms, some which could be domain-specialised, and others which could be domain-general.

1.1.1. Insights from case studies

To begin considering whether there are shared or separable neurocognitive mechanisms between language and music domains, it is useful to look at case studies. If language and music shared all neural networks, loss of musical ability would be an automatic consequence of loss of language ability and vice versa. Alternatively, if language and music shared no neural networks, there would be no cross-domain transfer of specific impairments between language and music domains.

Case studies show that language impairments do not necessarily lead to music impairments. Shebalin (1902–63) was an accomplished 20th-century composer who retained his musical abilities despite experiencing a substantial loss of language abilities after a left temporoparietal lesion caused by two strokes in his fifties (Luria et al., 1965). He had severe difficulties comprehending language and could not follow simple instructions like “point to your nose.” Nevertheless, he continued to analyse and correct his students’ compositions successfully and to compose orchestral works of high quality. In fact, Shostakovich (1906–75; another prominent Russian composer at the time) described Shebalin’s fifth symphony, written five years after his second stroke, as “a brilliant creative work, filled with highest emotions […] optimistic and full of life” (Mazzucchi et al., 2017). A similar case is that of MM, a 74-year-old professional composer who, after a stroke that caused enlargement of the sulci of the two temporal lobes, was unable to provide the names of animals and other non-musical objects. Still, he could name musical instruments (Tzortzis et al., 2000).

Since Shebalin and MM were professional musicians and had spent many years in musical training, they might have developed neural networks devoted to music that are not present in the general population. However, there is also evidence of musical abilities remaining intact despite the loss of language abilities in people who are not professional musicians. Patient NS was not a trained musician and had limited musical experience. After a right temporal stroke, Patient NS's aphasia meant that he could not understand the meaning of lyrics or recognise voices. However, he could distinguish between musical instruments based on their sounds and recognise melodies (Mendez, 2001).

Further case studies demonstrate a double dissociation between language and music abilities, signifying neural, developmental and evolutionary independence of functions (Plaut, 1995). As well as music abilities remaining intact despite the loss of language abilities, language abilities can also remain unchanged despite the loss of musical skills. For example, Patient HJ developed severe amusia after a stroke, which meant that he could identify choruses from songs by their lyrics but could not identify popular melodies that did not have associations with lyrics (Wilson et al., 2002). Meanwhile, his abilities for language and reasoning remained unaffected. Similarly, Patient CN suffered bilateral temporal lobe damage at the age of 35 that caused amusia. She could not identify or experience a sense of familiarity with music that was once very familiar to her. However, she recognised their lyrics and could identify non-musical sounds, such as animal noises (Peretz, 1996). Additionally, Peretz (1993) presents Patient GL's case, who had a specific deficit in implicit tonal knowledge. Patient GL could not recognise when a chord or tone did not fit within a harmonic context. This deficit affected his ability to recognise melodies and severely decreased his enjoyment of music. Nevertheless, he had no reported language deficits. Together, these case studies provide evidence of separable mechanisms for language and music processes.

On the other hand, other case studies reveal cross-domain deficits in language and music abilities, potentially indicating neurocognitive links between domains. For example, Hofman et al. (1993) present the case of a 73–year–old amateur musician who had an ischemic stroke in the lateral part of the parietal-occipital region of the left hemisphere. In his speech, he would get stuck and repeat particular words and sentences. Similarly, in his musical productions, he would get stuck on certain melodies and phrases. Midorikawa et al. (2003) report the case of a patient with Wernicke’s aphasia, a syndrome that creates difficulties in understanding language, who also had deficits in understanding rhythm in musical contexts. Furthermore, Peretz et al. (1997) present the case of 40-year-old Patient IR who had damage to her temporal lobes and right frontal lobe after surgery, leading to short-term and long-term memory deficits in language processing. At the same time, in the music domain, she could no longer memorise novel melodies and could not identify melodies that were once very familiar to her. Together, these case studies suggest that there could be shared mechanisms for language and music processes.

1.1.2. Insights from cross-domain expertise transfer effects

In addition to case studies, another type of evidence that informs the debate is the cross-domain transfer of expertise effects. Multiple studies report associations between musical training and improvements in language performance. Such studies could indicate shared neurocognitive mechanisms between language and music domains. Even without formal training, musical aptitude positively correlates with improved phonological awareness and better reading skills (Anvari et al., 2002). In healthy childhood development, active musical training appears to improve literacy acquisition, phonological skills, speech perception and speech segmentation (Besson et al., 2011; Butzlaff, 2000; Chobert et al., 2012; Degé & Schwarzer, 2011; François et al., 2013; Gordon, Fehd, et al., 2015; Kraus et al., 2014; Moritz et al., 2013; Saffran et al., 1996). Musical training also appears to aid second language learning (Lowe, 1998; Patscheke et al., 2016; Talamini et al., 2018). For example, professional musicians are faster at learning artificial languages than non-musicians (Dittinger et al., 2017; Francois & Schön, 2011). Based on such findings, second language learning tools are increasingly incorporating music (Sundberg & Cardoso, 2019).

Similarly, musical therapies have been designed to improve language skills (Albert et al., 1973; Vines et al., 2011). For example, melodic intonation therapy has been found to be an effective intervention in Broca's aphasia, a disorder of language production (Albert et al., 1973). In melodic intonation therapy, patients are taught to sing short phrases such as "How are you today?" The musical elements are gradually removed, and the patient's speech improves. Musical activities have been found to promote language and speech processing in several groups, including in older adults who do not have dementia (Fu et al., 2018), as well as those who do have dementia (Vink et al., 2013), stroke patients (Grau-Sánchez et al., 2018), children with autistic spectrum disorder (Wan et al., 2010) and children with reading difficulties (Cogo-Moreira et al., 2013).

A limitation in interpreting cross-domain expertise effects is that most experiments in this field are cross-sectional. Therefore, results might be explained by mediating factors, e.g. people who engage in musical training may have a predisposed higher motivation to learn. Yet, studies also observe cross-over effects when the provision of musical training is experimentally manipulated. Using a longitudinal design, Moreno et al. (2011) conducted a randomised-controlled experiment in which children received computer-based musical training or computer-based painting training. Post-intervention, children in the musical training condition had significantly improved reading performance than children in the painting condition. This result shows two crucial things. First, children's initial motivation to participate in musical training did not explain the improvement (the children did not choose whether they took part in musical training or painting training). Second, it suggests something special about the relationship between language and music, compared to the relationship between language and painting. In a similar study, Tierney et al. (2013) concluded that high school students had enhanced neural processing of speech after two years of group-based musical training than peers in a fitness-based training programme. Longitudinal studies in this field are unfortunately few, as providing musical training is time-consuming and expensive. Having said this, all together, the current evidence compellingly suggests that musical training can improve performance in the language domain.

The association between musical training and language improvements could be explained by shared neurocognitive mechanisms between the two domains. Patel (2011) proposes the OPERA hypothesis that provides a theoretical framework to explain why musical training might benefit language functions. It is an acronym, standing for **O**verlap (anatomical overlap), **P**recision (music relies more on the regions of anatomical overlap than language), **E**motion (musical training is associated with positive emotions that can reinforce learning), **R**epetition (musical training involves repetition) and **A**ttention (musical training activities usually require focused attention). Patel (2011) proposes that it is through these conditions that musical training may also be beneficial for language skills. If some neurocognitive substrates involved in music processing are required in both language and music domains, this could explain why musical training might lead to more efficient language processing. Therefore, cross-domain expertise effects could indicate shared mechanisms between language and music domains.

Identifying which processes might share neurocognitive substrates in language and music domains could guide effective musical interventions for language skills. In turn, this could lead to improved quality of life, learning and general cognitive functioning, and could protect against cognitive decline (Bird et al., 2019; Blom et al., 2017; Digard et al., 2020; Habib & Besson, 2009; Haukedal et al., 2018; Hilari et al., 2015; Markham et al., 2009).

1.2. Expectancy violations

In the present thesis, expectancy violations are used to investigate which processes might rely on shared or separable neurocognitive substrates between the language and music domains. Through everyday exposure to language and music, individuals learn their rules implicitly and form implicit expectations for events to follow these rules. Violations of these rules result in low-probability sentences or musical phrases, and individuals experience these as expectancy violations. Expectancy violations have been widely studied in the language domain and, to a lesser extent, in the music domain using brain imaging methods with high temporal resolution such as event-related potentials (ERPs). However, very few studies have made active comparisons of ERP effects between the two domains. Such comparisons could provide novel insights into what processes rely on shared or separable neurocognitive substrates between domains.

ERPs can be detected at the scalp and calculated from EEG (electroencephalogram) data. ERPs are the average of EEG waves that are time-locked to specific events, such as the onset of a stimulus or response. A benefit of the EEG method is that it has a high temporal resolution, in the order of milliseconds, which means that ERPs can show electrical activity associated with neurocognitive processes in real-time. This high temporal resolution is beneficial when studying fast neurocognitive responses, such as those to expectancy violations. ERPs are relatively small and occur within a large amount of noise. However, by averaging measurements over several trials, it is possible to increase the signal-to-noise ratio and identify robust ERPs. These ERPs can be used to compare neurocognitive events between different tasks or stimuli presentations, such as expected and unexpected events.

It is important to note the limitations of the ERP methodology. The inverse problem means that similar ERP effects do not unambiguously reflect the same neurocognitive substrates (Poldrack, 2006; Tarantola, 2004). EEG has a low spatial resolution. Electrical potentials are measured at the scalp, and the number, location and magnitude of the generating sources in the brain are unknown. There is a mathematically infinite number of solutions to this inverse problem. Therefore, ERP effects with similar scalp topographies do not necessarily suggest that there are shared underlying neurocognitive substrates. However, with targeted stimulus and task designs, differences in scalp topography, latency, and amplitude of ERP effects can together shed light on differences in neurocognitive substrates (Luck, 2014).

The current literature hosts a fierce debate about which processes in language and music domains might rely on domain-specialised or domain-general neurocognitive substrates (Koelsch, 2011a; Patel, 2008). Amongst the leading candidates are semantics, grammar, harmony and meter. These processes are the focus of this thesis. The following sections introduce their current positions within the debate, focusing on the ERP correlates of their expectancy violations.

Other processes are involved in the debate, but they are not the focus of the current thesis. One of these processes is pitch. It is perhaps more suitable to compare pitch processing between language and music domains using tonal languages, such as

Mandarin, to facilitate cognitive rather than only acoustic comparisons between domains, see Gandour (2012). The experiments presented in this thesis present the language stimuli in the English language. Likewise, comparisons between phonemes and timbres are a valuable part of the debate. Koelsch (2011a) points out that there is no clear difference in the definition of the words “phoneme” (used in language contexts) and “timbre” (used in music), and suggests that these acoustic features may be directly comparable between language and music domains. Although a valuable direction for future study, phonemes and timbres were not the focus of the current thesis due to their primarily auditory nature.

1.2.1. Expectancy violations in language

1.2.1.1. Semantic expectancy violations

First, semantic processing, and typical ERP effects associated with semantic ERP effects are introduced. In this thesis, semantics is defined in terms of meaning. Some words have similar meanings. For example, the words “common”, “usual”, “ordinary”, “familiar”, and “regular” have similar meanings. The meaning of a word can be inappropriate in some contexts. Therefore, it is possible to violate semantic expectancy. For example, the sentence “*The pizza was too hot to eat*.” is semantically expected. On the other hand, in the sentence “*The pizza was too hot to cry*.”, the final word is a semantic violation (Kutas & Hillyard, 1983). Methods of creating semantic expectancy violations are discussed in more detail in section 2.1.1.

Kutas and Hillyard (1980) conducted the classic semantic expectancy violation ERP study. They presented three types of sentences which varied in end type: a standard end type, e.g., “*It was his first day at work*.”, a semantic expectancy violation end type, e.g., “*He spread the warm bread with socks*.”, and a physical violation end type, e.g., “*She put on her high-heeled SHOES*.” They predicted that both semantic and physical violations would elicit a P300–type ERP effect, as such a response had been found for unexpected letters in words in a previous study by Shelburne (1972). As predicted, the physical violation words elicited a P300–type response (an increase in amplitude positivity peaking around 300 ms for the physical violations compared to the standard end type). However, to their surprise, the semantic violations elicited an N400 effect – increased negativity around 400 ms after stimulus onset for semantic violations compared to the standard sentences (Kutas & Hillyard, 1980). Since this classic study, several studies have investigated ERP effects associated with semantic violations. Together, they demonstrate that semantic violations robustly elicit an N400 effect (Gutierrez et al., 2012; Kamp et al., 2015; Kutas & Federmeier, 2010; Kutas & Hillyard, 1983; Moreno & Vázquez, 2011; Nigam et al., 1992; Rommers et al., 2013; Tiedt et al., 2020).

The N400 effect is characterised by a centroparietal increase in negative amplitude elicited by semantic violations compared to semantically congruent stimuli. It peaks around 400 ms after the onset of semantic violations (Kutas & Federmeier, 2010). More severe semantic violations elicit a larger N400 effect than less severe violations. For example, Kutas and Hillyard (1983) report that the N400 effect was larger for more severe semantic violations, e.g., “*The pizza was too hot to* <u>*cry*</u>*.”* compared to less severe violations, e.g., “*The pizza was too hot to* <u>*drink*</u>.”. The finding that the N400 occurs, but weaker, for less severe semantic violations supports the hypothesis that it is elicited by neurocognitive substrates that are specialised for semantic processing, rather than general expectancy violation processing.

The N400 occurs for various presentation types of semantic violations. For example, it occurs for semantic violations presented in American Sign Language (Gutierrez et al., 2012) and those that are semantically incongruent relative to their local contexts, such as neutral emotion words presented within either positive or negative word lists (Kamp et al., 2015). When the context makes semantic violations less crucial for stimulus interpretation, the amplitude of the N400 can be reduced (Rommers et al., 2013). Generally, reduced attention can reduce the amplitude of the N400 (Erlbeck et al., 2014). However, warnings that semantic violations are about to occur do not appear to affect the amplitude of the N400 (Moreno & Vázquez, 2011). Additionally, there is some evidence that language expertise may affect the N400 effect. Anurova and

Immonen (2017) report that there were larger N400 amplitudes for semantic violations in native speakers compared to non-native speakers.

The N400 is a robust ERP effect elicited by semantic violations, motivating research to investigate whether it can shed light on whether neurocognitive substrates for semantic processing are shared between language and music domains. This approach will be considered further in section 1.3.1.

1.2.1.2. Grammar expectancy violations

Grammar is another processes relevant to the debate between shared or separable neurocognitive mechanisms between language and music domains. It includes the rules with which types of words (e.g., tenses, verbs, nouns, singulars, plurals) and punctuation are used to form sentences. Each language has its own grammar rules. In English, "*The dog chased the cat*." is a sentence with acceptable grammar. On the other hand, "*Dog the cat the chased*" includes the same words but violates the grammar rules that English speakers have learned. Arguments for and against links between grammar and harmony processing are complex and are discussed in more detail in section 1.3.2. First, in the current section, ERP effects associated with grammar expectancy violations are introduced.

A common way to create a grammar expectancy violation is to present a word that does not fit with the sentence structure, e.g., due to unexplained mixtures of tenses, a verb instead of a noun, or using a singular word where the sentence creates expectancy for a plural word. An example comes from Patel et al. (2008), "*The sailors call for the captain and demands a fine bottle of rum*.". The word "*demands*" is used (which would be correct if the verb were attached to a singular "*sailor*") instead of "*demand*" (which is the correct verb match for current sentence, with the plural "*sailors*"). Methods of creating grammar expectancy violations are discussed in more detail in section 2.1.1.

The P600 is the most commonly reported ERP effect associated with grammar ex violations (Gunter et al., 2000; Hagoort et al., 2003; Kaan et al., 2000; Liao et al., 2020; Mehravari et al., 2015; Osterhout & Holcomb, 1992). Ill-formed sentences provoke a P600 effect – an increased positive amplitude peaking at around 600 ms for grammar violations compared to grammatically correct stimuli (Mancini et al., 2011)(Osterhout & Holcomb, 1992). The P600 tends to be more pronounced over the scalp's central or posterior areas (Palolahti et al., 2005).

The P600 is elicited by different types of sentence structure violations (Kaan et al., 2000; Loerts et al., 2013; Patel et al., 1998). For example, missing out crucial words, as in "*The sheep should grazing in the pasture*.", compared to "*The sheep should be grazing in the pasture*." (Mehravari et al., 2015). It is a reliable ERP effect for grammar violations, although some studies suggest it is not robust. For example, Hahne et al. (2012) found that while healthy control participants elicited the P600 effect for grammar violations, cochlear implant users did not. This finding could be interpreted alongside findings that suggest the P600 can be affected by language expertise. One study found larger P600 effects for grammar violations for people who were more proficient with a novel artificial language than those who were less proficient (Batterink & Neville, 2013). Furthermore, the P600 tends to be larger when the grammar violations are task-relevant, and it tends to decrease in amplitude or not to occur when grammar violations are task-irrelevant (Hahne & Friederici, 2002; Haupt et al., 2008; Lemhofer et al., 2020; Osterhout et al., 2002).

In addition to the P600, other ERP effects are associated with grammar violations in the language domain, although these are reported less often. These additional effects are typically either an early left anterior negativity (ELAN, 100–300 ms) (Molinaro et al., 2011; Palolahti et al., 2005) or late anterior negativity (LAN, 300–500 ms) (Friederici et al., 1993; Gunter et al., 2000; Hagoort et al., 2003) which sometimes accompany the P600. These earlier components are thought to be associated with automatic processing, while the P600 could indicate a cognitive process of grammar reanalysis (Hahne & Friederici, 1999).

Sometimes, there is also a “late positive potential (LPP)” reported for grammar violations (Baetens et al., 2011). At this point, it is apt to mention that the ERP literature is, at times, challenging to interpret due to its naming conventions. Different authors call similar components different names and some authors call components with different characteristics the same name. For example, a P600 could be a frontally distributed or posteriorly distributed effect. These differences in scalp topography may indicate different neurocognitive substrates. Still, the term “P600” gets assigned to both effects because of how ERP effects are named. In ERP naming practices, the “N” stands for “negativity”, and “P” stands for “positivity”. The polarity depends on the neurons' spatial arrangement that gives rise to the signal at that moment in time. The numbers following the “N” or “P” either denote the order of the components (e.g., the N2 follows the N1) or the time at which the effect peaks (usually given in ms, e.g., the N400 is a negative component, peaking around 400 ms). Earlier-onset components are usually named with the first (order) method and later-onset components with the second (ms). However, some components do not follow either of these naming methods, such as the LPP. The common characteristics of these components are even vaguer, allowing for increased flexible use of their terms. This flexibility leads to difficulties in cross-study interpretations. For example, in the grammar violation literature, it is unclear whether some reported LPPs would be named P600s if different authors presented them. An example of the challenges brought by ERP naming practices is that Patel et al. (1998) present a “P600” peaking around 1100 ms after the onset of grammar violations – it is unclear whether other authors would also call this effect a “P600”, or an “LPP”, or something else.

1.2.1.3. Comparing semantic and grammar processing

This thesis's primary focus is to investigate evidence of shared or separable neurocognitive substrates between language and music domains. To do so, it was also necessary to consider overlaps in semantic and grammar processing. If these processes rely on distinct neurocognitive mechanisms, then to make valuable conclusions about whether there are domain-specialised or domain-general mechanisms in language and music domains, it would be essential to compare each individually to processes in the music domain.

Hauser et al. (2002) propose a Faculty of Language model, suggesting a clear separation between the "faculty of language in the broad sense" (involved in grammar processing), and the "faculty of language in the narrow sense" (involved in semantic processing). Behavioural evidence further supports the separation of semantic and grammar processing. For example, while Ottl et al. (2017) predicted that including semantic information would facilitate learning the grammar of an artificial language, their two experiments showed no difference in grammar learning performance between with-semantics and without–semantics conditions.

In non-artificial languages, comparisons between semantic and grammar processes are less straightforward, as they are typically interlinked in the interpretation of a sentence. This idea can be demonstrated by expectancy violations, as grammar violations often also contain a semantic violation. For example, if the expected word is a noun, but participants receive a verb or an adjective (perhaps intended to be a grammar violation), the target contains both a grammar violation and a semantic violation. Therefore, sentences that experimenters intend to be a grammar violation may be interpreted as a semantic violation, and vice versa.

Consistent with this idea, Kim and Osterhout (2005) found a "semantic P600" effect in sentences that had correct grammar but were semantically obscure. For example, "*The hearty meal was devouring the kids."* As the P600 effect is typically associated with grammar violations, it could be interpreted that this "semantic P600" suggests a link between semantic and grammar processing. However, it is perhaps more likely that these sentences were processed as grammar violations. In the example given, the word "*devouring"* could be exchanged for the word "*devoured"*, which could be simpler (and therefore be processed with less effort) than switching the subject-object order ("*The kids were devouring the hearty meal*"). This finding demonstrates the difficulties of creating semantic and grammar violation stimuli. Crucially, it also supports the notion of separable neurocognitive substrates for semantic and grammar processing, as there was no N400 effect elicited by the sentences. This finding suggests that sentences with ambiguous semantics and grammar might be processed either as a grammar violation (as it seems has occurred in this case, reflected by the typical ERP effect for grammar violations - a P600 effect), or a semantic violation (which one would predict to be accompanied by an N400 effect).

Languages with gender-based articles for words provide a unique opportunity to separate semantic and grammar content. A gender article can be presented that is either congruent or incongruent with a target word (a manipulation of grammar), before a target word that is either semantically congruent or incongruent with the rest of the sentence. For example, the grammatical correctness of the gender article changes between “*la escoba”* (expected, meaning “the broom” in English) and *“el escoba”* (unexpected). However, by itself, the word “*escoba*” is semantically the same in both presentations (Wicha et al., 2004). Studies using such paradigms report that both the N400 and P600 occur, and have increased amplitudes for simultaneously presented semantic and grammar violations (Gunter et al., 2000; Loerts et al., 2013; Wicha et al., 2004). Such findings have been interpreted to suggest overlap in the neurocognitive substrates underlying semantic and grammar processing. However, these increased amplitudes could also be explained by other factors, such as increased attention to the target words. This alternative interpretation is supported, as other studies find that both N400s and P600s are larger when attention is drawn to the violations (Erlbeck et al., 2014; Hahne & Friederici, 2002; Haupt et al., 2008; Lemhofer et al., 2020).

Currently, perhaps the most compelling evidence of distinct neurocognitive substrates involved in semantic and grammar processing is that the main ERP effects associated with them, the N400 for semantic violations and the P600 for grammar violations, are different. Chapter 3 of the current thesis aims to test the case for their distinctness further. As the current evidence suggests separable semantic and grammar processing mechanisms, these processes are separately compared with harmony and meter processing in this thesis.

1.2.2. Expectancy violations in music

1.2.2.1. Harmonic expectancy violations

While semantics and grammar are critical processes in the language domain, harmony is an integral process in the music domain. In this thesis, harmony is defined in terms of its organisation in harmonic progressions, following Western traditional music theory rules. Patient GL, mentioned in section 1.1.1, could not recognise when a chord or tone did not fit within a harmonic context, but they had no language deficits (Peretz, 1993). This case study suggests that there are domain-specialised neurocognitive mechanisms for harmony processing. While there is no direct counterpart to harmony in language, it is possible that its processing shares some neurocognitive substrates with grammar, due to similarities in their hierarchical structures and rules. There is evidence for and against harmony processing sharing neurocognitive mechanisms with semantic and grammar processing. This evidence is discussed later in this chapter – see subsection 1.3.1 and subsection 1.3.2. Before that, harmonic expectancy violations, and their typical ERP correlates, are introduced.

Harmonic expectancies are learned implicitly through music listening (Bigand & Poulin-Charronnat, 2006; Ettlinger et al., 2011). Indeed, one study investigated whether children aged between 6- and 11-years-old from France, Australia, and Canada who had not received formal musical training had an implicit understanding of harmony. The children responded to whether a target chord was “good” or “bad” within a chord progression (Schellenberg et al., 2005). They responded faster and more consistently said that it was “good” when the target chord was a tonic (I) chord (which is the root of the key) compared to when the target chord was any other type of chord. This result suggests that, through passive music listening, Western children have implicit knowledge for Western tonal harmony, even without formal musical training. The implicit learning of harmonic rules is perhaps comparable (although not strictly equivalent) to grammar rules in language. While native speakers might not be able to explain grammar rules explicitly, they have an implicit understanding of them and know when they are being used incorrectly.

To describe how harmonic expectancy violations are created in Western tonal music, it is first necessary to cover some fundamental aspects of Western music theory. Traditional Western music has 12 standard notes (see Figure 1.1). These 12 notes are distinguishable by differences in pitch (frequencies of sound).

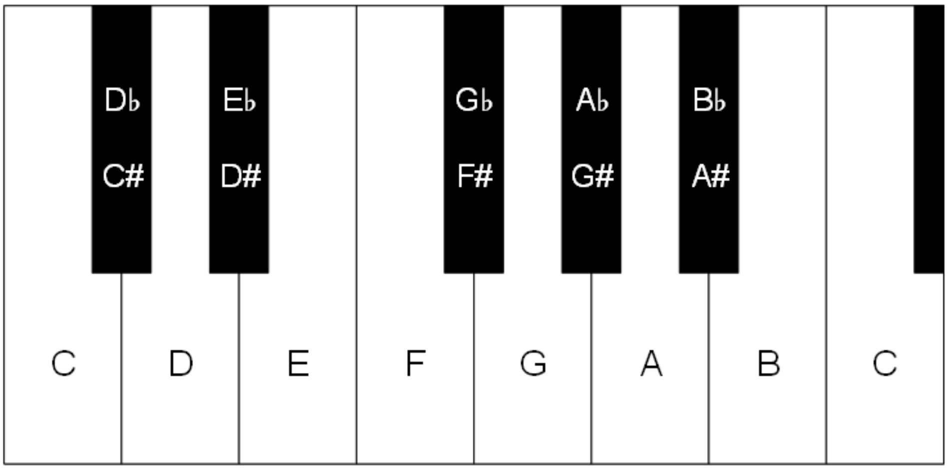


Figure 1.1 – The twelve notes in Western music, here demonstrated on a keyboard.

The distance between two adjacent notes (e.g., C and C#, or E and F) is a semitone, and the distance between two notes that are two notes apart is a tone (e.g., C and D, or E and F#). Each of the twelve notes has a major and a minor scale (which are set sequences of tones and semitones), which make up its key signature. Any of the 12 notes can be used as the first note of a major or minor scale. Often in Western tonal music, pieces or phrases of music are based on the notes of a chosen scale. These are referred to as "key signatures". For example, if a piece of music is based on the notes of *C major* and has a tonal root of *C*, it is in the key signature of *C major*. Each key signature includes seven triads (chords made up of three notes), one starting on each note of the scale. Each triad involves the root note and, relative to that root note, the third and fifth notes of the scale. These triads are commonly referred to in Roman numerals (I–viiº [1]), see Figure 1.2.

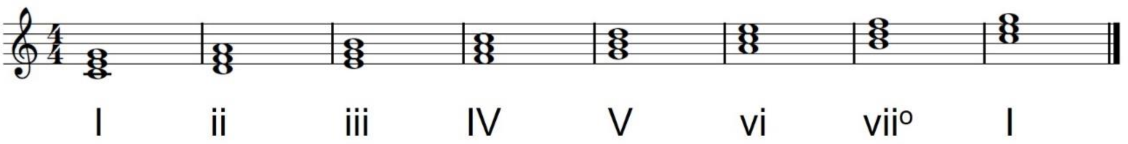


Figure 1.2 – The triads of *C* **I**–viiº**.**

The rules of Western music theory provide methods to create harmonic expectancy violations. In Western tonal music, some chord progression endings are more common (and therefore more expected) than others (Rohrmeier & Cross, 2008). The most common, and consequently the most expected, ending in Western tonal music is the perfect cadence. A perfect cadence is often found at the end of a piece or phrase and involves the dominant (V) chord with the dominant (V) note played in the bass, followed by the tonic (I) chord with the tonic (I) note played in the bass, see Figure 1.3. At the end of a phrase, the dominant (V) chord creates an expectation of the tonic (I). A harmonic expectancy violation can occur when a different chord follows the dominant (V) at the end of a progression. Ways to create harmonic expectancy violations are discussed in more detail in section 2.1.2.

---

[1] The ᵒ denotes that viiᵒ is a diminished chord. Diminished chords are minor triads with a lowered fifth (e.g., C, E♭, G♭).

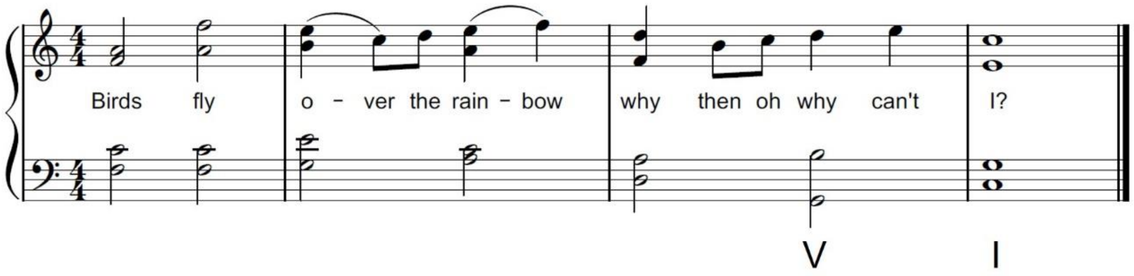


Figure 1.3 – An example of a perfect cadence (V-I) in *C major* shown alongside the famous melody of *Somewhere Over the Rainbow* (Arlen & Harburg, 1939). The scoring is original, created for this example. It is hoped that this example helps the reader to imagine the sound of a perfect cadence and its feeling of being expected.

The most commonly reported ERP effects for these types of harmonic violations are P300 effects (Koelsch, 2011a). In the context of harmonic violations, P300 effects tend to be characterised by an increase in positive amplitude peaking at around 300 ms after harmonic violations compared to harmonically expected stimuli (Beisteiner et al., 1999; Carrión & Bly, 2008; Janata, 1995; Loehr et al., 2013; Pei et al., 2004). P300 effects are more likely to occur when the task makes the harmonic violations explicit (Friedman et al., 2001).

Interpreting the P300 literature for harmonic violations is challenging, primarily due to ERP component naming conventions. First, P300 effects are sometimes, although not always, separated into “P3a” and the classic P300 or “P3b” (James et al., 2017; Janata, 1995; Polich, 2007; Steinbeis et al., 2006). The P3a is linked to working memory and attention, and the classic P300 (or P3b, from now on in this thesis referred to as a P300) is related to long term memory processing (Polich, 2007). In the context of harmony processing, the P3a is associated with dissonant chords presented individually, and it decreases in size with greater exposure to them. In contrast, the classic P300 is associated with incongruent harmony within a local context, and its characteristics are not affected by repeated presentations (Polich, 2007). The P3a and classic P300 do not tend to co-occur and appear to indicate different neurocognitive substrates.

Second, P300 effects are not only found for harmonic violations. They also occur for other types of expectancy violation, such as physical incongruities, sometimes termed “oddball” stimuli (e.g., a word presented in capital letters (Kutas & Hillyard, 1980), uncommon words (Polich & Donchin, 1988) and lies (Gao et al., 2014). While distinctions are not clear between P300 effects elicited by different types of stimuli, they appear to indicate different neurocognitive substrates as they have different scalp topographies. To improve the literature's clarity, direct comparisons of the characteristics of different P300 effects would be a valuable direction for future research.

Two primary forms of evidence suggest that P300 effects elicited by harmonic violations indicate neurocognitive substrates specific to harmony processing. First, these P300 effects can be larger in people with more exposure to music. For example, Steinbeis et al. (2006) found that musicians and non-musicians both elicited a P300 in response to unexpected harmonies, but this was considerably larger in the musicians. Second, P300s are sometimes larger with increased harmonic violation severity but still occur at lower severity levels. Janata (1995) presented chord progressions in major keys. The target chord was either the anticipated tonic (I) chord, a minor chord (a harmonic violation) or a dissonant chord (a more severe harmonic violation). The P300 effect was larger for dissonant chords than minor chords. This result provides evidence that the neurocognitive substrates indicated by the P300 effect are associated with harmonic processing, rather than general expectancy violation processing, e.g. auditory deviants. Again, to aid distinctions between ERP effects of similar amplitude direction and temporal characteristics, the names given to ERP effects should perhaps include information about the scalp distribution and the type of stimuli that elicit them. For example, “a P300 effect with a posterior scalp distribution elicited by lies”.

While the P300 is the most commonly reported ERP effect for harmonic expectancy violations, other ERP effects are reported for specific types. An N500 effect tends to be elicited by chords that create harmonic tension and suspense, such as Neopolitan sixths (Jentschke et al., 2014; Koelsch, 2011a; Koelsch et al., 2000; Loui et al., 2005; Steinbeis & Koelsch, 2008; Zhang et al., 2019). A Neopolitan sixth is the first inversion of a major chord on the flattened supertonic (the supertonic is chord ii). They are consonant and serve a unique function for modulation. They are not commonly heard but are popular in some composers' music, such as Scarlatti, Paisiello and Beethoven. The N500 tends to have a frontal distribution and is often accompanied by an ERAN (early right anterior negativity), peaking around 150 ms (Koelsch et al., 2000).

Crucially, the P300, N500, and ERAN effects appear to reflect cognitive processing, rather than merely sensory, processing (Collins et al., 2014). Simple auditory deviants, such as a sudden loud tone, elicit early ERP responses, such as a mismatch negativity effect (MMN), but not these later-onset effects (Koelsch, 2011a).

Like the P300, other neural correlates of harmonic violations appear to be affected by varied levels of exposure to music. Studies have reported a larger late positive potential (Jaśkiewicz et al., 2016) and a larger EANm2 (which is the MEG, magnetoencephalogram, equivalent of the ERAN) for musicians compared to non-musicians (Kim et al., 2011). Furthermore, a recent study compared the harmonic violation ERP effects of Chinese participants who were categorised into three groups based on their Western music proficiency: high, medium and low (Ma et al., 2018). The high proficiency group, who had formal Western musical training, elicited the N500 and ERAN ERP effects in response to the chords that created harmonic suspense, comparable to those in people who grew up exposed to Western music. However, the low and middle proficiency groups did not show any ERP responses for the target chords. Therefore, there is evidence that increased exposure to Western music through musical training strengthens harmonic expectations which, in turn, affects the neural mechanisms involved in harmonic processing.

But, could ERP effects correlated with harmonic expectancy violations provide evidence of separable neurocognitive substrates between language and music domains? In the next sections, the cases for and against domain-specialised neurocognitive substrates for harmony processing are discussed.

1.3. Cross-domain comparisons of neurocognitive substrates involved in language and music processing

1.3.1. Comparing semantic and harmony processing

There is debate about the extent to which semantics are comparable between language and music domains. Some authors suggest that emotional responses to music indicate musical meanings (Carr, 2004), and Koelsch (2011b) suggests that musical extracts can convey meanings such as “happy” and “light.” In support of this idea, Krumhansl (1997) found that participants reliably associated selected musical excerpts with either sadness, happiness, or fear. However, while music can perhaps convey some meanings, it does not seem to carry its own. This idea is reflected in that many scholars refer to “musico-semantics” or “extra-musical meanings”, which are musical features that provide metaphors for non-musical semantic meanings (Antovic, 2009; Meyer, 1956). As a simple example, the start- and end-points of a section of music could be a metaphor for the start and end-points of an event (Johnson & Larson, 2003).

In line with the idea that music does not have its own semantic meanings, a model proposed by Brown et al. (2006) suggests that language has distinct mechanisms from music for informational processes, such as semantics. Further supporting evidence comes from a case study of an amateur musician who had severe semantic dementia, who clearly demonstrated musical knowledge. He could sightread, make appropriate stylistic embellishments, and showed a demonstrable understanding of musical structures (Weinstein et al., 2011). All things considered, it seems likely that neurocognitive mechanisms involved in semantic processing might not be involved in the independent processing of music.

While attempts to find evidence of semantics specific to music could add valuable contributions to the debate, the current thesis focuses on the possibility that semantic processing shares neurocognitive substrates with harmony processing. There have been several attempts, using various methods, to test this. First, fMRI studies have attempted to localise semantic and harmony processing to similar brain regions. For example, one study found that both types of processes are correlated with spontaneous brain activity in the supramarginal gyrus and left superior temporal gyrus (Yu et al., 2017). The authors suggest that this result supports the hypothesis that language and music share neurocognitive substrates involved in semantic and harmony processes. However, fMRI findings cannot straightforwardly provide strong evidence for shared neural networks between language and music domains. There are multiple networks within each brain region, and these are difficult to tell apart with fMRI. This limitation of the fMRI method is clearly demonstrated in a study from a different field, showing that mirror neurons are interspersed with purely motor-related neurons in the pre-motor areas of the macaque monkey cortex (Rizzolatti & Craighero, 2004). While these neurons are involved in distinct mechanisms, they create comparable patterns of activation on fMRI scans. Therefore, such fMRI evidence cannot easily provide strong evidence of shared neurocognitive substrates for semantic and harmony processing.

Second, interactions in ERP effects for simultaneously presented semantic and musical stimuli have been interpreted to indicate shared neurocognitive substrates between language and music domains. Recent studies find that the N400 effect can be smaller (Du et al., 2020) and delayed (Calma-Roddin & Drury, 2020) when semantic violations presented simultaneously with background music. The authors suggest this could provide evidence of increased semantic integration difficulty due to shared neurocognitive resources between language and music domains. Furthermore, there have been reports of interactions in the N400 for simultaneously presented semantic and harmonic violations. Koelsch (2011b) found a reduced N400 effect when semantic violations were presented simultaneously with chords that created harmonic tension. This interaction could suggest shared resources for semantic processing and harmony processing. However, the N500 effect for harmonic tensions was not affected by simultaneous presentation with semantic violations, weakening the case for shared neurocognitive substrates. An alternative explanation is that, in these studies, the N400 amplitude was affected by diverted attentional resources (and it could, therefore, also be affected by non-musical stimuli, e.g., background noise) rather than shared neurocognitive substrates for language and music processing. This explanation is supported, as other studies show that the N400 amplitude can be reduced with reduced attention (Erlbeck et al., 2014). Further doubt over shared neurocognitive resources between the N400 and N500 comes from comparing their scalp topographies which, respectively, tend to be centroparietal and frontal and thus are distinctly different (Koelsch, 2005).

Third, studies using priming paradigms have attempted to shed light on shared neurocognitive substrates for semantic and harmony processing. A priming study found that a similar N400 effect was elicited when a target word was preceded either by semantically unrelated sentences or musical primes, compared to when they were preceded either by semantically related sentences or musical primes (Koelsch et al., 2004). Examples of the stimuli include priming the words "wideness" (rather than narrowness) with a musical extract with expansive harmonies, and the word "narrowness" (rather than wideness) with a chromatic and dissonant musical extract. Source analysis showed that the source of the N400 showed no difference in location for language and music primes. The authors suggest these results could indicate a neurocognitive overlap between language and music domains in terms of semantic processing. However, in line with the idea of "extra-musical meanings" (Antovic, 2009; Johnson & Larson, 2003; Meyer, 1956), it is possible that harmonies could provide metaphors that prime semantic processing in the language domain without there being shared neurocognitive substrates involved in processing in each domain. To investigate further, future research could investigate whether musical stimuli, primed with incongruent semantic words, also elicit an N400 effect.

Fourth, cross-domain expertise effects could indicate shared neural resources between language and music domains. Investigations of cross-domain expertise effects from musical expertise to semantic processing are few in number, but some notable findings exist. Rosslau et al. (2016) found that singers had increased late neuronal activity in the right temporal and left parietal areas than actors for both sung and spoken semantic violations. It is possible that, in line with the OPERA hypothesis (Patel, 2011), through their musical training, the singers had built up more efficient semantic processing mechanisms compared to the actors, who are also language experts. Moreover, Dittinger et al. (2017) found that in children with musical training, the N400 to unexpected newly-learned words from an artificial language was significantly larger than in children without musical training. One explanation of this finding could be that the children with musical training had improved semantic processing mechanisms, due to shared neurocognitive substrates with harmony processing, in line with the OPERA hypothesis (Patel, 2011). However, another explanation of this effect might be that the children with musical training built stronger expectancies for the words compared to children without musical training. This interpretation is supported as the children with musical training also performed better on a word recognition task, designed to test how well they had learned the artificial language. Currently, these cross-domain expertise effects provide only weak support for the hypothesis that semantic and harmony processes rely on shared neurocognitive mechanisms. Further research, controlling for factors such as language proficiency and general cognitive ability, is needed. Altogether, the current evidence in support of shared neurocognitive mechanisms for semantic and harmony processing is weak.

Other evidence provides support for the separability of semantic and harmony processing. Kunert et al. (2016) found no difference in participants' reaction times or closure judgements of harmonic stimuli when presented simultaneously with semantic violations. Additionally, Besson et al. (1998) asked musicians to listen to excerpts from operas sung acapella (without accompaniment). The target word was either semantically expected or unexpected and sung in or out of key. There were distinct ERP components for violations in the lyrics (N400) and the melodies (P300), and their expectedness did not appear to interact within these components. Moreover, the N400 effect is a robust effect for semantic violations in not only the language domain but also in other contexts, e.g. for smells (Sarfarazi et al., 1999) and line drawings (Nigam et al., 1992). However, despite several attempts, no study has reported an N400 effect for musical expectancy violations (Besson & Faïta, 1995; Besson & Macar, 1987; Miranda & Ullman, 2007; Paller et al., 1992). Thus, while the neurocognitive substrates indicated by the N400 appear to be specific to semantic processing in multiple presentations, they still do not appear to be involved in the music domain.

On balance, the evidence currently suggests that the neurocognitive substrates involved in semantic and harmony processes are likely to be separable. Chapter 3 of this thesis further investigates whether distinct ERP effects are associated with semantic and harmonic expectancy violations.

1.3.2. Comparing grammar and harmony processing

The case for shared neurocognitive substrates between grammar and harmony processing is more compelling than the case for overlaps in semantic and harmony processing. This debate is current and rapidly expanding.

First, there is some evidence to support the hypothesis that neurocognitive substrates involved in grammar and harmony processing are separable. There are specific aspects of grammar that do not have musical counterparts, such as nouns and verbs. Additionally, recently, Faroqi-Shah et al. (2020) investigated language and music processing in people who had developed agrammatic aphasia after a left hemispheric stroke. The results of a computer-based task showed that there were no differences in musical structure processing for agrammatic aphasics relative to neurotypical controls. The authors suggest this finding supports that, at least to an extent, there are separable neurocognitive substrates for structural processing in language and music structures.

Having said this, much of the literature focuses on drawing links between grammar and harmony processing. The theorising is mainly based on the fact that both grammar and harmony involve integrating smaller elements into a hierarchical system as sequences are processed over time. Thanks to grammar and harmony, in both language and music domains, phrases create structures that go beyond the sum of their parts, unlike most animal vocal communication systems (Hauser et al., 2002). Therefore, grammar and harmony might be fundamental parts of what makes language and music distinctly human (Nettl, 2000).

Different methods have been adopted to investigate links between grammar and harmony processing. First, the evidence suggests that both are localised to Broca's area (Cheung et al., 2018; Koelsch et al., 2005; Kunert et al., 2015; Maess et al., 2001; Patel et al., 2008; Seger et al., 2013; Sluming et al., 2002). This evidence makes sense, as Broca's area is thought to have a general role in processing and integrating sequential information over time (Tillmann et al., 2006). However, this evidence does not necessarily suggest shared neurocognitive substrates between grammar and harmony processing, as there can be several sub-mechanisms within one brain region, as discussed in section 1.3.1 (Rizzolatti & Craighero, 2004).

Second, apparent interactions between grammar and harmony processing may suggest shared neurocognitive resources. Patel (2008) proposed a shared syntactic integration resource hypothesis (SSIRH) that suggests that language and music share neurocognitive substrates for structural processing. There have been direct attempts to find empirical support for this hypothesis. For example, studies have presented harmonic expectancy violations simultaneously with semantic and grammar errors. Results suggest that sentence closure judgements are less accurate (Kunert et al., 2016) and reading times are longer (Slevc et al., 2009) when harmonic violations are presented simultaneously with grammar violations but not when they are presented simultaneously with semantic violations. These studies could support two hypotheses: first, that semantic and harmonic processes rely on different neurocognitive substrates, and second that grammar and harmony might rely on shared ones. However, Perruchet and Poulin-Charronnat (2013) point out that the allocation of attentional resources could explain these interactions posited to support the SSIRH (Slevc et al., 2009). Grammar violations tend to cause a greater re-evaluation of the sentence than semantic errors, which requires more attention. The simultaneous presentation of harmonic violations is likely to divert attention from grammar processing. Therefore these findings might be explained by neurocognitive substrates involved in attention allocation rather than suggesting shared substrates that are specialised for grammar and harmony processing.

Third, a P600 effect has previously been reported to be elicited by harmonic violations. However, it appears that this effect has only been reported for a specific paradigm. The classic study by Patel et al. (1998) was the first to compare ERP effects associated with grammar and harmonic violations directly in a within-subjects design. They report a P600 effect for both grammar and harmonic stimuli that were difficult to integrate into the sentences or chord progressions. This effect was statistically indistinguishable between the two types of expectancy violation, in ANOVAs carried out at midline and lateral electrodes between 500–800 ms. However, the stimulus design complicates interpretations. Both grammar and harmonic expectancy violations were presented in the middle of the sentence or chord progression. In the language task, the target for ERP analysis (at 0 ms) was the onset of the word after the grammar incongruity. For example, in the sentence "*Some of the senators endorsed the* promoted *an old idea of justice*.", the word "promoted" was the word that was difficult to integrate into the sentence, and the word "an" was the target word for analysis. Similarly, in the music task, the target of ERP analysis (the 0 ms point) was the onset of the chord presented after the harmonic expectancy violation chord (which was an out-of-key chord). Therefore, it is not clear that these are P600 effects, in the same sense as other reported P600 effects, that characteristically occur 600 ms after the onset of the violation.

Furthermore, in Patel's (1998) study, the language stimuli were presented in the timing of 4.4 syllables per second (or one syllable per 227 ms), while the chords were presented 500 ms apart. This design meant that there was a difference in the time interval between violations and targets between language and music tasks. In the language task, the temporal distance between the grammar expectancy violation and 600 ms (the typical P600 peak) was 827 ms (227 ms + 600 ms). However, they report that their effect was maximal between 800–900 ms after the word following the grammar violation, which is between 1027–1127 ms (227 + 800 to 900 ms) relative to the grammar violation. In the same way, for the music stimuli, the "P600" effect peaked around 1100 ms (500 ms + 600 ms) after the onset of the harmonic expectancy violation. The difference in temporal presentations between tasks, alongside the fact that the target stimuli for ERP analysis were the word or chord after the violation, affects the ability to compare ERP effects directly between language and music tasks, and calls into question whether these effects should be considered “P600 effects”.

Perhaps these effects could instead be considered to be P800 effects. Zioga et al. (2016) report a P800 effect related to both prosodic and harmonic expectancy violations between 850–1200 ms after violation onset. In their study, they suggest that the P800 is “amodal”, and associated with the auditory modality, rather than it being specific to the language or music domains. Patel et al. (1998) presented both language and music stimuli in the auditory modality. Therefore, it is possible that Patel et al.’s (1998) P600 effect might be better considered a type of P800 effect, which are not thought to be specifically associated with grammar or harmony processing, but a more general auditory re-evaluation effect. It is necessary to test further whether P600 effects might occur for harmonic violations and, if so, whether they could indicate shared neurocognitive substrates with grammar processing.

Fourth, there could be similarities between the main ERP effect associated with grammar violations (the P600 effect) and the main ERP effect for harmonic violations (the P300 effect). There are for P300 effects elicited by non-musical stimuli (Bornkessel-Schlesewsky et al., 2011; Coulson et al., 1998; Münte et al., 1998; Sassenhagen et al., 2014), and there are some similarities between P600 and P300 effects. Both are often affected by task saliency (Hahne & Friederici, 2002; Osterhout et al., 1996) and do not occur if violations are not detected (Batterink & Neville, 2013; Sassenhagen & Fiebach, 2019). Differences in latencies between the P600 and P300 do not necessarily indicate different neurocognitive substrates, as it is possible that these latencies simply depend on the complexity of the stimulus, or other steps of processing. A recent study found that it was possible to decode grammar violation trials that elicited a P600 with multivariate pattern analysis classifiers trained on oddball trials that elicited a P300. These classifiers were just as accurate at decoding P600 trials as classifiers trained on the P600 data itself (Sassenhagen & Fiebach, 2019). In further support of the hypothesis that these P600 and P300 effects have shared neurocognitive substrates, both tend to have centroparietal scalp topographies (Sassenhagen & Fiebach, 2019). Moreover, both

P600s and P300s have been reported for other types of violations which supports the possibility that they are general indicators of structural processing, and could rely on shared neurocognitive substrates. P600 effects have been reported in non-language contexts for example in numerical sequences (Núñez-Peña & Honrubia-Serrano, 2004), spelling errors (van de Meerendonk et al., 2011), and double-checking processes (Kolk & Chwilla, 2007). Similarily, P300 effects have been elicited by several types of violations, including physical incongruities, sometimes termed “oddball” stimuli, e.g., a word being presented in capital letters (Kutas & Hillyard, 1980), uncommon words (Polich & Donchin, 1988) and lies (Gao et al., 2014). Therefore, P600 and P300 effects might both be elicited by general structural integration processes, only differing in temporal characteristics due to the complexity of the stimulus or other steps of processing.

Surprisingly, previous studies do not seem to have directly compared P600 effects for grammar processing with P300 effects for harmony processing despite the debate about shared neurocognitive substrates for grammar and harmony processing. The P300 effects that have previously been compared with P600 effects in the literature are within language contexts (Frisch et al., 2003; Osterhout, 1999). As mentioned previously, naming conventions in the ERP literature do not allow for a straightforward distinction between different types of P300 effects, elicited by different types of stimuli. Consequently, it remains unclear whether P300 effects elicited by harmonic expectancy violations might share neurocognitive substrates with P600 effects elicited by grammar violations.

The debate about whether grammar and harmony share neurocognitive substrates persists. Patel et al. (1998) provide motivation to test further whether harmonic violations elicit a P600 effect and, if so, whether that P600 effect is comparable to those elicited by grammar violations. Additionally, there is a gap in the literature for direct comparisons between P600 effects for grammar violations and P300 effects for harmonic violations. Chapter 3 in the current thesis aims to investigate these questions further.

1.3.3. Meter expectancy violations

While semantic, grammar and harmony processing might rely on distinct neurocognitive mechanisms, the literature currently suggests that meter processing might rely on domain-general ones. In this section, the roles of meter in language and music domains are introduced, the evidence for and against domain-general meter processing mechanisms is presented and ERP effects for meter expectancy violations are discussed. Then, interactions between meter processing and semantic, grammar and harmony processing are considered.

To define meter, it is perhaps useful to distinguish it from rhythm. While rhythm involves irregular patterns of emphasised and non-emphasised beats over time; regularly repeated beats characterise meter (Lerdahl & Jackendoff, 1983). Meter is an essential feature in both language and music domains. In the music domain, the meter provides temporal regularity against which all other details of a piece of music are projected. Meter is organised by the regular grouping of beats into bars, and time signatures show how many beats are in each bar. The grouping of musical meter into bars means that there is typically anticipation for certain types of events occurring on certain beats (Benjamin, 1984). For example, a 4/4 meter is four regular beats to a bar with a strong beat falling on the first beat of every bar. Meter provides crucial contributions to the temporal organisation of music, and it aids musical perception and interpretation (Palmer & Krumhansl, 1990).

Meter's importance to the music domain is further reflected as musical training, even for just one year, improves children's ability to tap to a beat (Slater et al., 2013), and adults with formal musical training perform better than adults without formal musical training on tasks where they are required to detect changes in meter (Yates et al., 2017). Sensitivity to different types of meter differs depending on the prominent meter present in the music an individual listens to. For example, in Turkish music, meter, or "usul", is organised in rhythmic patterns with more or less complex inner structures of beats of differing duration and emphasis. Turkish listeners are more sensitive to changes in usul compared to American listeners (Yates et al., 2017). Additionally, Zhao et al. (2017) found that musicians had larger MMNs to deviant meters compared to non-musicians and suggest this reflects that musicians build more accurate and efficient neural mechanisms for meter processing through musical training.

Meter is, perhaps, less straightforward to conceptualise in the language domain. However, evidence shows that it plays a crucial part in language comprehension. As in music, in the language domain, meter can be thought of as regular beats in time. Cutler's rhythmic segmentation hypothesis proposes that listeners use the form of meter prevalent in their language as a prelexical cue to word boundaries, e.g., the beat in English, syllable in French or mora in Japanese (Cutler & Otake, 2002). Developing an understanding of meter is vital for language development (Suppanen et al., 2019). Meter helps people to learn vocabulary (Jusczyk et al., 1999), separate words (Mattys & Samuel, 1997) and interpret the grammar of sentences (Schmidt-Kassow & Kotz, 2008). Furthermore, one study found a rhythmic priming effect on spoken sentences – when the priming rhythm matched the rhythm of the sentence, phoneme detection was enhanced (Cason et al., 2015).

As might be predicted based on the crucial role of meter in language processing, there are strong relationships between meter and language abilities. In one study, sensitivity to meter predicted phonological awareness and reading development, accounting for over 60% of the variance in reading ability, when age and IQ were controlled for (Huss et al., 2011). Additionally, a recent study found that meter and rhythm perception and production were strong predictors of phonological awareness in 3- to 4-year-old children. (Politimou et al., 2019). Furthermore, temporal perception, auditory rhythmic perception and tapping to a beat are all performed significantly more poorly by children with specific language impairments, such as dyslexia, compared to children without language impairments (Corriveau & Goswami, 2008; Richardson et al., 2004; Sallat & Jentschke, 2015; Wolff, 2002). This dysfunction in meter processing could contribute to the atypical development of phonological representations for spoken words which is the primary cognitive characteristic of dyslexia across languages (Leong & Goswami, 2014).

Similar to how music experts appear to have enhanced meter processing, language training might enhance meter processing. In one study, bilingual participants performed better on a meter change sensitivity task than monolinguals (Kalender et al., 2013). In another study, simultaneous language interpreters were more accurate than people who were not language experts when deciding whether piano melodies were rhythmically the same or different (Elmer et al., 2010).

Although meter is not as regular in spoken language as in the music domain (Jackendoff & Lerdahl, 2006), there are similarities between them. For example, like dancing or finger tapping along with musical beats, hand gestures that accompany speech often line up with strong stresses in the speech (McNeill, 1994). Additionally, an important cue for language interpretation is a “heavy” syllable, such as one with a long vowel, or that closes with a consonant (Spencer, 1995). This could be similar to how there are often longer and louder notes on strong musical beats (Jackendoff & Lerdahl, 2006). Moreover, in poetry, the meter is often grouped into regular beats, as it is in music (Obermeier et al., 2013).

There are further striking links between rhythm and meter in language and music domains. Jusczyk et al. (1999) propose that in respect to meter, at least during infancy, the human brain does not treat language and music as strictly separate domains. Additionally, Patel and Daniele (2003) compared the instrumental music of several composers from England and France, including Elgar, Vaughan Williams, and Holst (English) and Ravel, Debussy and Poulenc (French). They found that English music had more stressed–unstressed rhythm pairs than French music, as is characteristic of their spoken languages.

The neurocognitive substrates associated with meter processing might align with Dynamic Attending Theory (Jones & Boltz, 1989). According to this theory, neural oscillations synchronize with regular beats, enhancing temporal expectancies for upcoming events. In line with this theory, Port (2003) found that there were regular neural pulses elicited by words spoken in English that followed a regular meter pattern (as though spoken in time with a metronome). They conclude that, during everyday speech, these metrical fields are implicit, and speakers control the degree to which they allow them to constrain the timing of their speech. Listeners use meter to interpret speech, as it increases attention towards stressed word syllables (Pitt & Samuel, 1990). Likewise, in music, listeners use meter to anticipate and interpret the music (Benjamin, 1984).

Few studies directly investigate links between meter processes in language and music domains. As mentioned above, Midorikawa et al. (2003) report the case of a patient with Wernicke’s aphasia, which involves difficulties in understanding language, who also had deficits in understanding and following the rhythm and meter of music. Furthermore, there are links between musical training and improved beat and meter perception in people with Williams syndrome (for which musical interest is a prominent aspect of the phenotype) (Lense & Dykens, 2016). In the same study, those with higher meter perception skills also had higher adaptive communication skills. Additionally, domain-general mechanisms involved in meter processing could perhaps explain the case of the 73–year–old amateur musician, introduced in section 1.1.1. The amateur musician would get stuck and repeat words and sentences as well as melodies and phrases (Hofman et al., 1993), perhaps due to a deficit involving meter processing.

The idea that language and music might share neurocognitive mechanisms for meter processing is not without debate. Vuust and Witek (2014) show that meter in the music domain stimulates audio-motor pathways, more so than in the language domain, which the authors suggest is due to links between meter in music and movement (Hickok & Poeppel, 2007). Additionally, Jackendoff (2009) argues that linguistic intonation contours are specific to the language domain, while the typical grouping of meter into beats in a bar is unique to the music domain. Having said this, Jackendoff (2009) still concludes that an underlying sense of meter might be domain-general.

People develop expectations for metric patterns based on temporal regularity and the metric patterns of previous events (Tillmann & Lebrun-Guillaud, 2006). Therefore, meter violations can be created by violating temporal regularity. Despite the the debate about whether language and music share mechanisms for meter processing, ERP effects for meter violations have not been directly compared between domains. Most of the existing literature investigating meter violations focuses on meter violations in the music domain.

There are two main, and related, ways of creating meter violations in the music domain. The first is to present stimuli that are off the main beat or in other words, where the interstimulus interval is varied. These irregular interstimulus intervals tend to produce an N1 response (Davis et al., 1966; Foldal et al., 2020). For example, Fitzroy and Sanders (2015) found a frontally distributed N1 effect that was more negative for strong beats (on the beat) compared to weak beats (off the beat). In another study, an irregular interstimulus interval onset of tones produced a frontal N1 and posterior P2 effect compared to a regular interstimulus onset (Menceloglu et al., 2020). Furthermore, Besson and Faïta (1995) found that the frontal N1–P2 peak-to-peak amplitude was larger when the final musical notes in a familiar melody were delayed by 600 ms compared to notes that were presented at the anticipated times. This effect was larger for familiar melodies, where expectancy might have been stronger, compared to unfamiliar melodies. The second main method used to create meter violations is to present a "silent beat" compared to a sounded beat. For example, Raij et al. (1997) showed that the frontal N1 and posterior P2 amplitudes were both larger to tones played after a silent beat compared to tones played after a continuous sequence of tones played on the beat. In this method, the sequence of tones sounding on the regular beat creates an expectation of a continuation of the sequence. In contrast, the silent beat caused a break in the regular sequence. This break might lead the listener to expect that the sounded beats had ended, so a following beat might be processed as a meter violation.

ERP effects for meter violations can be moderated by musical expertise. Jongsma et al. (2005) found that both N1 and P2 effects elicited by a sounded beat following a silent beat, compared to a sounded beat following a sequence of other sounded beats were larger in musicians compared to non-musicians. Additionally, Habibi et al. (2014) found that the N1 effects to a delayed note were significantly larger for musicians compared to non-musicians. In another study, the N1 amplitude was more negative for meter violations after participants had been trained on a sensorimotor rhythm task with regular meter (Kober et al., 2015). This finding further suggests that the N1 amplitude is driven by learning and expectations. Perhaps differences in meter violation ERP effects between musicians and non-musicians could be explained by the musicians directing more attention to irregular meter stimuli, as they could have an enhanced ability to follow regular meters through their musical training.

Few studies have attempted to investigate ERP effects associated with meter violations in the language domain. In one study, metrically unpredictable words elicited a more negative N1 amplitude than metrically predictable words (Cox et al., 2016). In that study, four words were presented and the fourth (target) word either had the same "Expected" stress pattern (trochaic or iambic) as the previous three words or the opposite "Unexpected" stress pattern[2]. Another study found that inconsistent stress patterns in rhyming couples that were read were associated with an increase in amplitude negativity between 80–155 ms (an N1–like effect), compared to consistent stress patterns (Breen et al., 2019). In another study, rhythmic mismatches between target words and primes presented in the auditory modality elicited an N1 effect (Zhang & Zhang, 2019). Therefore, N1 effects could be reliable ERP effects associated with meter violations in the language domain, for stimuli presented in both visual and auditory modalities.

Together, the results of different studies appear to suggest that ERP effects point towards domain-general neurocognitive substrates for meter processing. The apparent existence of an N1 effect for meter violations in language and music domains (Besson et al., 1997; Breen et al., 2019; Cox et al., 2016; Fitzroy & Sanders, 2015; Raij et al., 1997; Zhang & Zhang, 2019) suggests that there could be domain-general

---

[2] In a trochaic stress pattern, the first syllable is stressed, and the second syllable is unstressed. In an iambic stress pattern, the first syllable is unstressed, and the second syllable is stressed.

neurocognitive substrates. There is no previous evidence of a P2 effect for meter violations in the language domain, as there is in the music domain (Menceloglu et al., 2020; Raij et al., 1997), but this could be due to differences in how the meter violations have been created in language and music domains. The N1 and P2 are generally thought of as being attention-related effects (Luck, 2014) and, therefore, their potential associations with meter processing in both language and music domains could be consistent with the Dynamic Attending Theory (Jones & Boltz, 1989), which suggests that meter evolved to create readiness for, and therefore more efficient processing of, upcoming stimuli.

Despite the beat being the driving meter in the English language (Cutler & Otake, 2002; Pitt & Samuel, 1990; Port, 2003), and interstimulus intervals potentially having robust ERP correlates in the music domain (Besson & Faïta, 1995; Raij et al., 1997), no previous studies seem to have investigated beat violations in the language domain. This was the aim of Chapter 4 of the current thesis. Additionally, the question of whether there are cross-domain expertise effects on ERP effects for meter violations has not been investigated previously. If, for example, musical training moderates meter violation ERP effects in both language and music domains, then the case for domain-general neurocognitive substrates for meter processing could be strengthened.

1.3.4. Does meter processing rely on separable mechanisms from semantic, grammar and harmony processing?

After identifying that meter processing might share domain-general neurocognitive substrates, it is useful to consider whether meter processing is separable from potentially domain-specialised processes, such as semantics, grammar and harmony. If so, the hypothesis that there are distinct neurocognitive substrates associated with semantic, grammar, harmony and meter processing could be supported.

Some theories suggest that meter processes interact with both semantic and grammar processes. Gordon et al. (2015) found that rhythm perception accounted for 48% of the variance in grammar performance, after controlling for non-verbal IQ, socioeconomic status and prior musical activities. They suggest that in typical language comprehension, meter works in conjunction with grammar and semantics to allow the listener to predict when important parts of the speech signal are coming up. In line with this idea, some studies find interactions in ERP effects for semantic and grammar violations when stimuli are presented simultaneously with meter violations.

Findings are mixed for empirical studies investigating links between semantic and meter processing. Rothermich et al. (2010) found that the classic ERP effect for semantic violations (the N400) was significantly larger for metrically irregular sentences compared to metrically regular sentences. The meter violations were created by presenting final (target) word with the “Expected” stress pattern of syllables (trochaic or iambic), based on the previous words in the sentence or the opposite “Unexpected” stress pattern. In contrast, another study found that the N400 was smaller when semantic violations were presented with irregular meter compared to when they were presented with regular meter (Li et al., 2019). The authors of both studies interpret that the interaction in the N400 component could indicate shared neurocognitive substrates involved in meter and semantic processes. However, others find no interactions between meter and semantic expectancy violations in the N400 effect (Magne et al., 2007). Currently, the relationship between meter and semantic processing remains unclear.

There is also debate about whether there are shared neurocognitive substrates for grammar and meter processing. Another study found that listening to regular musical sequences before a grammar judgement improved performance of assessing grammatical correctness compared to rhythmically irregular musical sequences (Chern et al., 2018). Crucially, in that study, there was no effect of rhythm on two non-linguistic control tasks which could suggest a special overlap between rhythm and grammar mechanisms. A recent study showed participants rhythmic stimuli that were either regular or irregular before showing them sentences that were either grammatically expected or grammatically unexpected. They found that there were reduced P600 effects when participants had been primed with irregular, rather than regular rhythmic stimuli (Canette et al., 2020). On the contrary, Schmidt-Kassow and Kotz (2009) found that when grammar violations were presented simultaneously with interstimulus interval violations, there was a larger P600 effect, compared to when the two violation types were presented separately. Moreover, when analysed separately, there were similar P600 effects for both grammar and meter violations, perhaps providing further evidence of shared mechanisms. Altogether, the current evidence does not provide a clear view of whether there are shared or separable neurocognitive substrates for grammar and meter processing.

In the music domain, most of the literature investigates links between meter processing and pitch processing. Peretz and Coltheart (2003) proposed a model, based on Fodor's (1983) modularity of the brain approach, suggesting a modular organisation of music processing. In this model, they propose that pitch processing is separable from temporal processing (rhythm and meter). Previous empirical research supports the idea that meter processing is separable from other processes in the music domain such as meter, focussing on pitch. For example, one study reports that stroke patients with right hemisphere lesions can have severely disrupted melody processing but intact meter processing (Vignolo, 2003). People with congenital amusia have difficulties with pitch, but their rhythm abilities remain intact (Peretz & Hyde, 2003). For example, in one study, people with pitch deafness due to congenital amusia were able to interpret ambiguous drum rhythms and synchronise their movements to the beat of popular music (Phillips-Silver et al., 2013). Further findings indicate a double dissociation between language and music abilities. Case studies report that people can experience significant difficulties processing and interpreting rhythm and meter but can perform well on pitch-based tasks (Midorikawa et al., 2003; Peretz & Zatorre, 2005). Therefore, it is possible that meter processing relies on neurocognitive substrates that are separable from other processes in the music domain.

However, there is some contradictory evidence, suggesting links between pitch and meter processing, and it is currently unclear whether there are separations between harmony and meter processing. Pitch judgements (Jones et al., 2002) and melody completion judgements (Tillmann & Lebrun-Guillaud, 2006) have been found to be more accurate when the meter is regular, compared to when it is irregular, suggesting links between pitch and meter processing. In terms of harmony, Janata et al. (2002) found that temporal asynchrony judgements for a target chord were less accurate when the chord was preceded by harmonically unrelated chords compared to when the chord was preceded by chords in the same key (harmonically expected). Furthermore, harmonically expected target chords have been judged as being better fitting (Schmuckler & Boltz, 1994) and more complete (Tillmann & Lebrun-Guillaud, 2006) when they were played with a regular meter compared to an irregular meter. Likewise, Jung et al. (2015) found that metrically regular presentation of harmonic violations increased reaction times to their correctness judgements. An alternative explanation of these findings could be that attention distractions were compounded for unrelated harmony and irregular meter. In that case, these findings might not enable inferences about the neurocognitive substrates specifically involved in harmony and meter processing. Instead, there could be general attentional resources that are drawn on by meter as well as pitch and harmony. Therefore, the question of whether meter and harmony rely on shared neurocognitive mechanisms remains.

A recent novel study investigated ERP effects for simultaneous presentation of harmonic and meter violations. Zhang et al. (2019) presented chord progressions where the target chord either provided a sense of harmonic closure (the final two chords were V–I, a perfect cadence), or harmonic suspense (the final two chords were I–IV). These target chords were either presented in a regular meter relative to the rest of the chord progression (one chord every 600 ms) or slightly earlier (524 ms after the penultimate chord). The N500 effect, characterised by an increase in negativity at around 500 ms for the harmonic tension chords compared to tonic (I) chords, was only found when the meter was regular, not when the meter was irregular. The authors suggest that this result indicates shared neural substrates between meter and harmony. The N500 ERP effect appears to indicate different neurocognitive substrates to the P300 effect. Although both ERP effects tend to have frontal distributions when elicited by harmonic violations, the N500 occurs for chords that create harmonic tension while the P300 occurs for harmonically incongruent stimuli. Therefore, to strengthen understanding of shared or separable neurocognitive substrates between meter and harmony processing, it is necessary to investigate whether there are interactions between meter and harmony for harmonic violations that elicit the P300 effect.

In Chapter 5, this thesis tests for interactions in ERP effects for meter violations and semantic, grammar and harmonic violations. If there are no interactions in these ERP effects, it could support the hypothesis that there are some specialised neurocognitive substrates for semantic, grammar, harmony and meter processing.

1.4. This thesis: Rationale, layout and research questions

The question of the extent to which neurocognitive substrates for processes in language and music domains are shared or separable is undeniably complex. ERP studies provide promising insights, indicating robust ERP effects for semantic, grammar and harmony processing. The evidence suggests that ERP effects are separable for semantic and harmony processing (Besson & Faïta, 1995; Besson & Macar, 1987; Miranda & Ullman, 2007; Paller et al., 1992). However, those involved in harmony and grammar processing might overlap (Patel et al., 1998; Patel et al., 2008) which could, in turn, indicate overlap in the neurocognitive substrates upon which they rely.

Cross-study comparisons of ERP effects do not provide robust tests of distinct underlying neurocognitive substrates. In such comparisons, apparent differences in ERP effects could be due to differences in the stimulus or task design, acquisition methods or analysis protocols. With this in mind, in the current thesis, within-subject designs were favoured, stimuli were matched between language and music tasks where possible (e.g., in length and timing of presentation), task difficulty was monitored, and the scalp topography of ERP effects was presented in addition to their temporal characteristics. The semantic, grammar and harmonic violations were created in such a way that they were violations because they were not the strongly anticipated word or chord. Crucially, these methods allowed for active tests of whether ERP effects that occurred for one type of expectancy violation also occurred for another type. Furthermore, cluster-based permutation statistics were conducted in addition to the classic ANOVA methods. This enabled effects that were not predicted a-priori to be identified. Therefore, the limitations of ANOVA methods in detecting effects that might be comparable for different violation types were, at least partially, overcome. Together, these methods allow for stronger conclusions about whether ERP effects associated with semantic, grammar and harmony processing could contribute to the debate about the extent of shared and separable neurocognitive mechanisms between language and music domains.

Chapter 2 demonstrates how the stimuli were created for the experiments presented in this thesis. The aim was to create expectancy violations that allow for matched stimuli and task design, where possible, between language and music tasks. The stimuli were based on previous studies that elicit the N400 for semantic violations, the P600 for grammar violations and the P300 for harmonic violations. As the previous literature lacks systematic research investigating the relative expectancies of different harmonic violations, Experiment 1 was designed to test how “Expected”, or “Unexpected” participants rate different harmonic endings.

Chapter 3 presents two experiments (2 and 3). They test whether the semantic, grammar and harmonic expectancy violations, created in Chapter 2, elicit distinct ERP effects. To test whether any ERP effects are robust and whether they are task-independent, two tasks are presented: in one (Experiment 2), the violations are task-relevant, and in the other (Experiment 3), the violations are task-irrelevant.

Furthermore, several theories and empirical studies suggest that meter processing could be domain-general. Such theories appear to currently be tentatively supported by the ERP literature, as a few studies have reported meter violation ERP effects that appear to have similar characteristics in language and music domains. However, surprisingly, no previous study appears to have made active comparisons between the two domains. This literature gap motivates the current thesis to actively compare ERP effects for meter violations between language and music domains.

In Chapter 4, two experiments (4 and 5) are presented. Experiment 4 is a behavioural experiment, designed to create meter violation stimuli with comparable expectancy ratings between language and music domains. Experiment 5 is an ERP experiment investigating whether there are comparable ERP effects for meter violations In language and music domains.

Additionally, some previous findings, discussed in section 1.3.4, suggest that neurocognitive mechanisms for meter processing might interact with those for semantic, grammar and harmony processing. Motivated by these findings, the current thesis tests for interactions in ERP effects when meter violations are simultaneously presented with semantic, grammar and harmony violations. Chapter 5 includes two experiments (6 and 7). Experiment 6 focuses on expectancy violations in the language domain, and Experiment 7 focuses on the music domain. If there is evidence that there are interactions in ERP effects for simultaneous presentations, it could suggest that the neurocognitive substrates that they indicate are not specialised for one type of processing. In turn, this could complicate the interpretation of whether the ERP effects could indicate domain-specialised or domain-general neurocognitive mechanisms.

Finally, within- and cross-domain expertise effects were investigated in all experiments, where sample characteristics allowed, based on participants' English language experience and musical training. In line with the OPERA hypothesis (Patel, 2011) one might predict that there will be within-domain expertise effects and cross-domain transfer of expertise effects. Such effects could support the case for shared neurocognitive substrates between language and music domains. The main research questions were as follows:

- Is there evidence that ERP effects for semantic, grammar and harmonic expectancy violations indicate separable neurocognitive substrates? (Chapter 3)
- Are there similar ERP effects for meter violations in language and music domains? (Chapter 4)
- Do ERP effects for meter violations interact with those for semantic, grammar or harmonic violations? (Chapter 5)
- Are there cross-domain effects of language or musical expertise observable in any ERP effects? (Chapters 3, 4 and 5)

2. Chapter 2 – Creating expectancy violations

2.1. Introduction

To be able to investigate ERP effects associated with expectancy violations in language and music domains, it was first necessary to identify suitable stimuli. Chapter 2 shows how the semantic, grammar and harmonic expectancy violation stimuli were created for the subsequent experiments presented in this thesis. These stimuli were based on experiments that had reported the N400, P600 and P300 effects in previous literature because these ERP effects appear to be most robust effects for semantic, grammar and harmony processing (respectively), and so these were the effects of interest for this thesis.

The aim was to create stimuli with certain similarities between domains, e.g., in length, number and timing of presentation. For ERP studies, it is beneficial to present the target stimulus at the end of the trial so that it is possible to collect behavioural responses (e.g. task responses and reaction times) that can aid the interpretation of any ERP effects. Therefore, methods for creating violations were chosen where the expectancy violation would be presented at the end of each trial. Furthermore, all types of violation (semantic, grammar and harmonic) were violations because they are not the strongly anticipated word or chord. These similarities allow for attempts at analogous task paradigms between language and music domains, facilitating comparisons of expectancy violation ERP effects between domains.

Chapter 2 is in two parts: the first part focuses on how language-based expectancy violations were created (both semantic and grammar violations); and the second part focuses on creating harmonic expectancy violations, which includes an experiment designed to test the expectancy of harmonic expectancy violations systematically.

2.1.1. Creating semantic and grammar expectancy violations

Semantic expectancy violations can be created in various ways, by presenting a word that is semantically unexpected within the context of a sentence. A simple way to create a semantic violation is to write a factual statement, that is incorrect (Fischler et al., 1983). Semantic expectancy violations can also occur when the meaning of words does not fit with the rest of the sentence. This can be in the middle of a sentence, as in Patel (2011), "*Anne scratched her name with her tomato on the wooden door*." To investigate behavioural correlates of ERP effects, it is often preferable to present a violation at the end of a sequence. For example, "*Your accent is very attractive.*" (expected) versus "*Your accent is very yellow.*" (unexpected) or "*The make-up highlights your cheeks*." (expected) versus "*The make-up highlights your flowers."* (unexpected) (Moreno et al., 2016). A classic example of this comes from Kutas and Hillyard (1983), who presented different levels of semantic expectancy in their sentences, e.g., "*The pizza was too hot to eat.*" (expected), "*The pizza was too hot to drink."* (unexpected), and "*The pizza was too hot to cry.*" (very unexpected).

Grammar expectancy violations be created by presenting a word that does not fit within the grammatical context of the rest of the sentence. For example, grammar expectancy violations can be created simply by changing the tense of a word so that it is incongruent. In the sentence "*The patient met the doctor while the nurse […] show the chart during the meeting"*, where the word "*show" i*s used instead of the correct tense, "*showed*" (Gouvea et al., 2010). Another example is in Patel et al. (2008), "*The sailors call for the captain and demands a fine bottle of rum*.", where the word "demands" is used instead of "demand". Another way to create grammar expectancy violations is to miss out words that allow the sentence to make sense. For example, the sentence "*The sheep should grazing in the pasture*." is grammatically incorrect. It would be grammatically correct if the word "*be"* were placed immediately before "*grazing*" (Mehravari et al., 2015). In most of the previous literature, as illustrated in the examples given in this paragraph, grammar expectancy violations are presented in the middle of a sentence. For the current thesis, the aim was to create grammar expectancy violations that were comparable to semantic expectancy violations. To achieve this, each type of violation is presented in the final words of the sentences. This approach provides an opportunity to compare ERP effects for semantic and grammar expectancy violations, and to compare them both in an equivalent way to ERP effects for harmonic expectancy violations.

The ways of creating the stimuli were based on the same principles as semantic expectancy violations that had previously elicited the N400 effect and grammar expectancy violations that had previously elicited the P600 effect (Gouvea et al., 2010; Kutas & Hillyard, 1983). Semantic expectancy violations were created by presenting the final word of the sentence that had a meaning that was incongruent with the meaning of the rest of the sentence. Grammar expectancy violations were created by presenting word forms that were incongruent with the structure of the sentence – for example, unexplained mixtures of tenses, singular versus plural words, or verbs versus nouns. As the semantic and grammar expectancy violation stimuli were closely based on previous studies, it was not deemed necessary to conduct a separate experiment to test participants' expectancy ratings of the language stimuli at this stage of the thesis (this is tested explicitly in section 3.2.2 Experiment 2: Behaviour results).

As mentioned in Chapter 1, section 1.2.1.3, a complexity in creating semantic and grammar expectancy violations is that grammar violations are often also semantic violations, and semantic violations are often also grammar violations. For example, if the expected continuation is a noun, but participants receive a verb or adjective, the semantic content of the continuation will be unexpected in addition to its grammar content. Kim and Osterhout (2005) found that sentences containing a semantic expectancy violation could be processed as a grammar expectancy violation if the grammar expectancy violation were more salient than the semantic expectancy violation. Taking these findings into consideration, when creating the stimuli in the current thesis, it was aimed to make semantic and grammar expectancy violations that were unlikely to be interpreted as the other.

To achieve this, sentences with high cloze probability endings were selected for the current experiments. Previous studies show that when grammar expectancy violations are presented with standard probability words, it is more likely that they will be processed as a grammar expectancy violation, rather than a semantic expectancy violation (Gunter et al., 2000; Loerts et al., 2013). Therefore, for the grammar expectancy violations, the target word was kept close to the standard meaning, and only the word form was changed. For example where the standard sentence was "At night the elderly woman locks a <u>door</u>.", for the grammar expectancy violation sentence, "<u>doors</u>" was presented as the final word, which was grammatically unexpected as the preceding word "a" built anticipation of a singular word.

Cloze probability is the probability that a word is given as a sentence continuation in a cloze task in which participants are asked to guess the next word of a sentence. The sentences used in this thesis were adapted from sentences created by Block and Baldwin (2010). Block and Baldwin (2010) tested 498 sentences using a sample of *N* = 400 participants. Participants were shown sentence stems and asked to fill in what they thought the final word should be. For example, 99% of participants completed the sentence stem "*She could tell he was mad by the tone of his* ____." with the word "*voice"*. This sentence had high cloze probability. On the other hand, only 18% of participants completed the sentence stem "*After failing, he realised he needed a new* ____." with the word "*plan"*. This sentence had low cloze probability. The sentence stems selected for this thesis all had high cloze probability (> 90%). This cut-off is a stringent cut-off for a sentence to be considered as high-cloze probability, as was recommended by Bloom and Fischler (1980).

There is a further crucial benefit of using high-cloze word endings. As is discussed later in this chapter, to create harmonic violations, a different chord was presented in place of a highly anticipated tonic (I) chord. By using high-cloze probability words, an analogous design can be made between domains. The semantic and grammar violations can similarly be presented in place of the highly anticipated word. This enables more confident comparisons of the neurocognitive substrates indicated by any associated ERP effects.

The high cloze endings were used for the expected "standard word" end type. The related semantic, unrelated semantic and grammar expectancy violation end types were then created, based on the methods that have been used in the previous literature (Gouvea et al., 2010; Kutas & Hillyard, 1983). The related semantic violations were incongruent with the sentences but were semantically possible. It was deemed valuable to include this less severe related semantic violation condition, because if the same ERP correlates are found for these as for more severe semantic expectancy violations, the hypothesis that the correlates are domain-specialised for semantic processes, rather than general expectancy violation processes, will be supported. The unrelated semantic violations did not fit the sentences and were semantically impossible. The grammar expectancy violations were created by presenting a word that either caused an unexplained mixture of tenses, singular words versus plurals or verbs versus nouns. It was not clear how different levels of expectancy for grammar violations could be created, and no previous study appears to have attempted to do so. Therefore, while it would be theoretically valuable to include two levels of expectancy for grammar violations, as for semantic violations, only one level of grammar violations was included.

A full list of the 24 sentence stems and their four ending types can be found in Appendix A. Seven sentences were eight words in length, and eight sentences were seven words in length. The lengths of the sentences were matched to the lengths of the chord progressions, which are described in the next section. Table 2.1 shows one example of a sentence stem with the four sentence end types.

Table 2.1 – An example of one sentence stems with the four sentence end types

Sentence stem example	Sentence ending	Sentence end type
He cashed his new paycheck at the	bank.	Standard word
	school.	Related semantic
	elephant.	Unrelated semantic
	banked.	Grammar violation

2.1.2. Creating harmonic expectancy violations

Based on the principles of Western tonal music, harmonic expectancy violations can be created in many ways. The focus of the current thesis is harmonic violations that elicit a P300 effect, as these could enable insights into whether P300 effects for harmonic violations share similar characteristics, such as scalp topography, with P600 effects for grammar violations, which could indicate shared neurocognitive substrates, as have been suggested for other P300 and P600 effects (Sassenhagen & Fiebach, 2019; Sassenhagen et al., 2014).

To create a harmonic violation that elicits a P300 effect, one can present a chord that is not the anticipated chord at that particular stage within the harmonic progression. Therefore, it is first necessary to create strong anticipation for a different chord. The most common, and, therefore, the most expected, ending in Western tonal music is the perfect cadence. A perfect cadence is often found at the end of a piece or phrase and involves the dominant (V) chord in root position[3], followed by the tonic (I) chord in root position. The dominant (V) chord creates the expectation of the tonic (I), so harmonic expectancy is violated when a different chord follows the dominant (V) at the end of a progression. Therefore, one method to create harmonic expectancy violations is to build anticipation of a perfect cadence (V–I) but then end the chord progression on one of the seven triads of the key signature (ii–vii °) that is not the tonic (I). When a chord progression creates anticipation of the tonic (I), any other triad within the key (ii, iii, IV, V, vi, vii °) will be unexpected but is still somewhat harmonically fitting. Previous studies have used this method to create harmonic expectancy violations, and some report P300 effects (Janata, 1995; Steinbeis et al., 2006). However, there is no consensus on which triads create the most reliable harmonic expectancy violations. Different authors use different triads to create expectancy violations (see Table 2.2). In the past literature, when authors have

---

[3] Root position means that the lowest note of the chord is the key–note of the chord. For example, a C major chord (with notes C, E, G), is in root position if C is the lowest note of the chord. The alternatives are E being the lowest note (first position) or G being the lowest note (second position).

previously used more than one triad to create harmonic violations, they have not separately analysed the expectancy for each triad type. Therefore, it is currently unclear how unexpected people find these violations, relative to each other.

Table 2.2 – Triads used by previous authors in place of the tonic (I) following the dominant (V) to create harmonic expectancy violations.

Triads	Reference
mediant (iii)	Guo and Koelsch (2016)
subdominant (IV)	Poulin-Charronnat et al. (2006)
subdominant (IV), supertonic (ii)	James et al. (2008)
submediant (vi)	Janata (1995)
submediant (vi), subdominant (IV)	Steinbeis et al. (2006)
submediant (vi), supertonic (ii)	Kim et al. (2011)

2.2. **Experiment 1: Testing the expectancy of harmonic expectancy violations**

2.2.1. Experiment 1: Rationale and aims

Experiment 1 was designed to investigate the expectancy of different chord progression endings, measured by whether participants report that they are “Expected” or “Unexpected”. The aim was to identify chord endings that participants reported as mostly “Expected”, those that were mostly “Unexpected”, and those that had an even mixture of “Expected” and “Unexpected” responses. If this were achieved, later experiments could investigate potential differences in ERP effects between more severe and less severe harmonic expectancy violations. If similar ERP effects are found for less and more severe harmonic expectancy violations, the hypothesis that these ERP effects are domain-specialised for harmony processing, rather than being an indicator of general expectancy violation processing, would be supported. Additionally, Experiment 1 aimed to systematically evaluate the relative expectancies of different triadic (ii– vii ^{o}) endings, as previous studies use various combinations of these triadic endings as harmonic expectancy violations, without reference to a systematic evaluation of their expectancies, which limits the interpretation of any related findings.

Most people listen to music, regardless of whether they have had formal musical training or not. Therefore, the detection of harmonic expectancy violations should not necessarily require formal musical training (Bigand & Poulin-Charronnat, 2006). As expectations about harmony are formed through music listening, it might be predicted that the relative expectedness of the seven triadic endings might be associated with the frequencies that people have heard each triad (ii– vii °) in the music they listen to. It is not possible to keep track of the range and amount of music that an individual has listened to over their lifetime. As a result, it is not possible to estimate the empirical probabilities of expected chord transitions for individuals. The frequencies of harmonic chord progressions of Bach's chorales are thought to be approximations of those across Western tonal music and are thought to be comparably familiar to both expert and non-expert listeners (Trainor & Trehub, 1994). Therefore, several previous studies have used chord progressions from Bach's chorales to represent common chord progressions in Western tonal music (Jaśkiewicz et al., 2016; Koelsch, 2005; Steinbeis et al., 2006). With this in mind, harmonic progressions were adapted from Bach's chorales to create the musical stimuli in the current thesis (Bach-Gesellschaft, 1892).

As discussed above, a reliable way to create the expectancy of the tonic (I) is to use a perfect cadence (that is, a dominant (V) followed by a tonic (I) chord). The perfect cadence is very common in Western music, and this is empirically supported, as Rohrmeier and Cross (2008) found that the frequency of V–I progressions in Bach's chorales is high (75%). They reported that the order of the frequency of the other triads in the key following the dominant (V) is: submediant (vi; 11%), subdominant (IV; 5%), mediant (iii; 5%), supertonic (ii; 4%) and leading tone (vii °**;** 0%). Jonaitis and Saffran (2009) suggest that expectancies for harmony are learned through statistical frequencies, similar to how languages are learned. In light of this, for Experiment 1, it was predicted that the relative expectancies of the triadic endings would follow the order of frequenxy reported by Rohrmeier and Cross (2008).

Although forming harmonic expectations does not require formal musical training (Bigand & Poulin-Charronnat, 2006), those with musical training might be exposed to more music, and have an overt understanding of musical theory, and, therefore, develop stronger harmonic expectations than those without musical training. Indeed, previous research suggests that musicians, but not non-musicians, appear to experience a subdominant (IV) chord as an unexpected event (James et al., 2008; Poulin-Charronnat et al., 2006). Furthermore, James et al. (2017) report that musicians were able to detect subtle harmonic expectancy violations (for example, chord endings that ended on the first inversion of the chord, rather than in root position), whereas non-musicians were not. Understanding any differences in the perception of harmonic violations between amateur musicians and non-musicians will aid future interpretations of any differences in ERP correlates between these groups. Consequently, an additional aim of Experiment 1 was to investigate whether there are differences in how expected participants find different chord progression endings, based on their musical training experience.

2.2.2. Experiment 1: Method

Ethics

The University Teaching and Research Ethics Committee at The University of St Andrews approved this study, code PS13356 (see Appendix P).

Participants

Eighteen university students volunteered to participate in this study. The data of two participants were not included in the analysis. One participant was excluded because they responded before the question prompt on over 40% of the trials (only trials with responses given from 100–5000 ms after the onset of the question prompt were considered for analysis). The other participant was excluded because they are a professional musician, with over 40,000 hours of accumulated practice time, and this experiment aimed to focus on people with amateur levels of musical training and non-musicians.

The final *N* = 16 participants (2 male, Age: *M* = 23 years, *SD* = 7 years, range = 18–49 years). In the recruitment advertisement for this experiment, people were asked to sign up only if they considered themselves either a “musician” or a “non-musician”. Participants were categorised into amateur musician and non-musician groups based on their view of whether they were a “musician” or a “non-musician”[4]. There were seven amateur musicians and nine non-musicians. Descriptive statistics of participants’ musical training experience for amateur musicians and non-musicians are in Table 2.3. The frequency distribution of accumulated practice times is in Appendix F. Each testing session took 45 minutes, and participants were reimbursed £4.

Table 2.3 – Descriptive statistics of musical training experience for amateur musicians and non-musicians, Experiment 1.

	Amateur musicians (*N* = 7)		Non-musicians (*N* = 9)	
Measure	M (SD)	Range	M (SD)	Range
Number of years	11.0 (4.6)	5–18	3.4 (2.3)	1–7
Practice time (hours)	3294 (1388)	1844–5850	198 (120)	52–446
Overall ability rating (1–5)	3.3 (0.5)	3–4	2.3 (1.0)	1–4
Listening hours	26.3 (15.5)	10–56	7.8 (3.7)	2–14

Apparatus

Stimuli presentation and data collection were run on Experimental Run Time System Version 3.32 (Beringer, 1994). The stimuli were presented at a viewing distance of roughly 80cm on a 17–inch cathode ray tube (CRT) monitor. Responses were recorded using a keypad with two keys set 10cm horizontally apart.

---

[4] Group memberships matched with the cut–off chosen after collecting data from all participants, of > 1000 hours of practice time for participants to be categorised as an amateur musician, and < 1000 hours of practice time for non-musicians (see Appendix F).

Materials

Chord progressions

The chord progressions were developed from passages in Bach's chorales: BGA 39 (BWV 253–438) selection of 185 chorales (Bach-Gesellschaft, 1892). There were three criteria for inclusion: the passages were required to i) end on a perfect (V–I) cadence, ii) not modulate, and iii) include seven or eight chords. The first two criteria were chosen so that they created anticipation of the tonic (I) chord, and the third criterion was chosen so that the length of the chord progression stimuli would match the length of the sentence stimuli (discussed in section 2.1.1) to support attempts at analogous task design between language and music tasks in future studies. A total of 72 chord progressions were chosen and their harmonic progressions were identified and transcribed. For details about the original chord progressions, such as the specific bars of the chorales they were based on, see Appendix B.

In the main experiment, participants heard 216 chord progressions, all played by a piano sound. The 72 chord progressions were presented with three end types: tonic (I), related harmony, and unrelated harmony (see Figure 2.1). Due to the inclusion criteria mentioned above, the chord progression stems created anticipation of the tonic (I) chord. Therefore, the tonic (I) chord was the expected chord progression end type.

For the related harmony chords, chords that were within the same key of the rest of the chord progression but, critically, not the tonic (I) chord was used. The six other chords within the key were used for the related harmony end type: ii-vii °. Hence, the related harmony chords were not the expected outcome but are not severe harmonic expectancy violations, as they remain within the key signature. Each of these six chord types related to 12 chord progression stems. For all chord progression end types, those same 12 chord progression stems were presented in the 12 major keys. Seven were eight chords in length, and five were seven chords in length (to match the characteristics of the sentence stimuli, discussed in section 2.1.1).

For the unrelated harmony chords, triads that were not in the same key as the rest of the chord progression were created. These chords were created by moving all notes from the related harmony chords up or down a semitone (for each set of 12 chord progression stems, half were moved up and half were moved down). Therefore, in isolation, the unrelated harmony chords had the same harmonic characteristics as the related harmony chords (e.g., major, minor or diminished) but are more severe harmonic expectancy violations, as they do not belong in the harmonic context created by the previous chords.

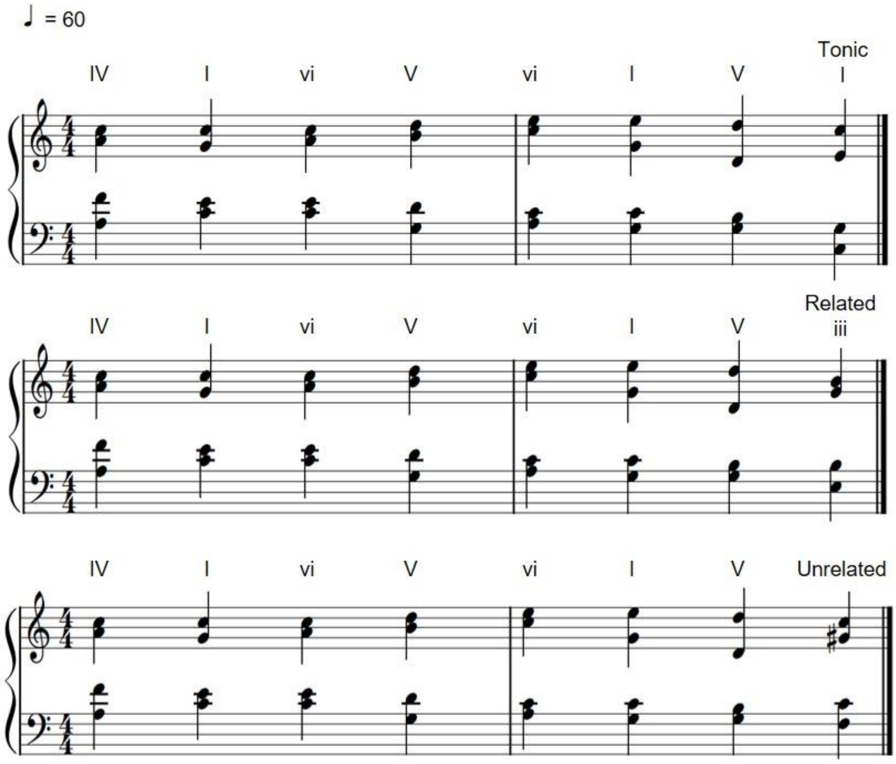



Figure 2.1 – One of the 72 chord progressions with the three chord progression end types: tonic (I), related harmony (iii in this example) and unrelated harmony (iii moved up one semitone in this example).

Musical Training Questionnaire A self-report demographic questionnaire developed by Jentzsch et al. (2014), see Appendix C. This questionnaire requires participants to record their accumulated practice hours for each instrument they have learned[5], the age they started and finished learning each instrument and the number of hours they spend listening to music per week. It also requires participants to rate their understanding of music history, music theory, music reading and overall music ability on the following scale: 1 = not able, 2 = limited, 3 = average, 4 = above average, 5 = very able.

Procedure

Participants sat by themselves in an isolated booth and listened to the chord progressions over loudspeakers. Participants were required to listen to the chord progressions and decide whether the final chord was expected or unexpected. For each chord progression, one chord was presented every 1000 ms (see Figure 2.2). 1000 ms after the final chord of each chord progression, a question mark appeared on the computer screen, and participants responded: "Expected" or "Unexpected" by pressing one of the two keys on the keypad. The assignment of the response buttons (left and right) was counterbalanced between participants.

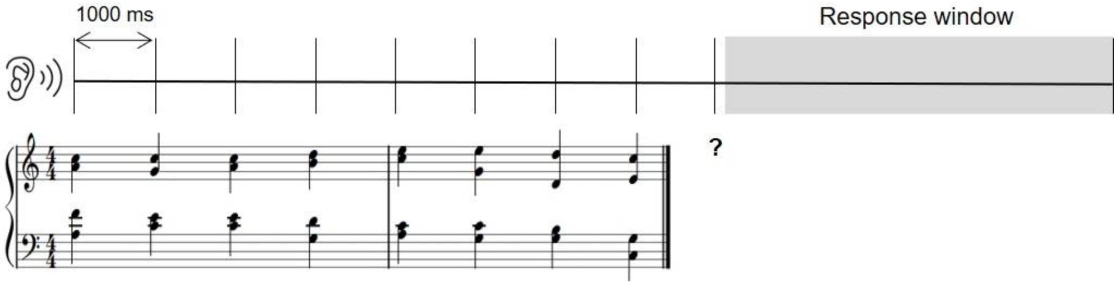



Figure 2.2 – The timing of stimuli presentation for the experimental task. One chord was presented every 1000 ms. The example in the illustration uses the tonic (I) chord ending. The accepted response window was between 100 ms and 5000 ms after question onset (shaded in grey).

Participants completed a practice trial of eight randomly chosen progressions to get used to the task. Then, the 216 experimental trials were presented in six blocks of 36 chord progressions. Participants were able to take breaks after each of the six blocks.

---

[5] Participants were asked to estimate hour many hours on average they spent practising each instrument per week for each year they had played it. These hours were then multiplied and added up to estimate overall accumulated practice time. This method was chosen in attempt to increase the accuracy of approximations of accumulated practice time.

When participants had completed the task, they were asked to fill in the Musical Training Questionnaire. It was deemed important that participants fill out the Musical Training Questionnaire after the task was completed, in case increasing the saliency of their musical training experience (or lack thereof) affected their responses.

Analysis

For all experiments presented in this thesis, SPSS v25.0 was used for behaviour analysis. The package ggplot2 was used for plotting bar graphs in R Studio (Wickham, 2009). The standard deviations and standard error bars are corrected for within-subjects designs. They were calculated using the summarySEwithin() function in the Rmisc package in R. This function removes inter-subject variability, calculating the within-subjects adjusted values as suggested by Morey (2008).

As is the case for all ANOVA analyses presented in this thesis, if the results of Mauchly's test of sphericity were significant, it is reported, and the degrees of freedom are corrected. If the Greenhouse-Geisser $\varepsilon < 0.75$, then degrees of freedom were corrected by the Greenhouse-Geisser method. On the other hand, if the Greenhouse-Geisser $\varepsilon > 0.75$, then degrees of freedom were corrected by the Huynh-Feldt method. For all ANOVAs, pairwise comparisons were corrected with the Bonferroni method.

Only trials with responses given from 100–5000 ms after the onset of the question prompt were considered for behaviour analysis. After the inclusion criterion was applied, the analysis included, on average, *M* = 210/216 trials (*SD* = 6.65) per participant. See Table 2.4 for summary statistics of how many trials out of a possible 72 per chord progression end type (tonic (I), related harmony and unrelated harmony) the analysis included.

Table 2.4 – Mean (*SD*) number of trials included in the analysis after the inclusion criteria were applied in Experiment 1. The maximum possible number of trials for each chord progression end type = 72.

Chord progression end type	Mean number of trials (*SD*)
Tonic (I)	69.8 (3.1)
Related harmony	70.9 (1.5)
Unrelated harmony	68.9 (3.3)

2.2.3. Experiment 1: Results

Response frequency

The percentage of trials to which the participants responded “Expected” was used to calculate the frequency of “Expected” responses, as percentages. With these percentages, two analyses were conducted.

The first analysis, a 3 x 2 mixed ANOVA, investigated whether chord progression end type (3 levels within-subject: tonic (I), related harmony and unrelated harmony) or musical training (2 levels between-subject: amateur musicians and non-musicians) affected how expected the final chords were (see Figure 2.3). Mauchly’s test indicated that sphericity had been violated χ^2 (2) = 8.32, *p* = .016. Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates (ε = .68). The results showed an effect of chord progression end type on the frequency of “Expected” responses *F*(1.36, 19.01) = 341.90, *p* < .001, ηp^2 = 0.96. All pairwise comparisons were significant at the *p* < .001 level with Bonferroni correction. As predicted, the tonic (I) chords received mostly “Expected” responses (*M* = 95%, *SD* = 1%), the related harmony chords received a mixture of responses (*M* = 46%, *SD* =4%) and the unrelated harmony chords received the least % of “Expected” responses (*M* = 8%, *SD* = 1%). There were no main effects or interactions associated with musical training (*p*s > .05).

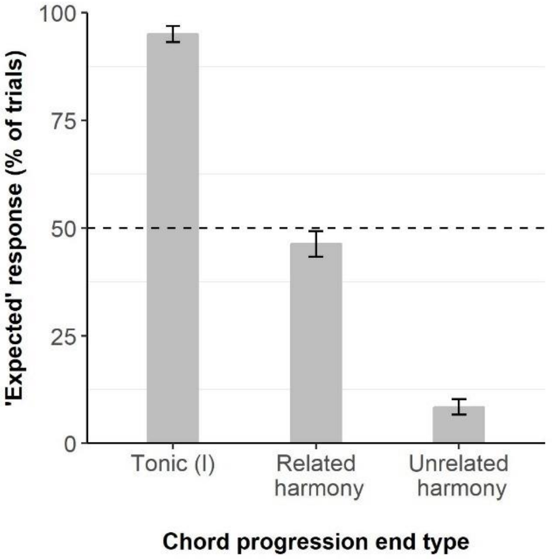


Figure 2.3 – Mean % of trials with an “Expected” response, for the three chord progression end types: tonic (I), related harmony and unrelated harmony. The dashed line at 50% illustrates the 50% (chance) level. Within-subject *SE* bars are included.

The second analysis, a 6 x 2 mixed ANOVA, investigated whether triad type in the related harmony chord type (6 levels within-subject: ii, iii, IV, V, vi, and vii^{o}) or musical training (2 levels between-subject: amateur musicians and non-musicians) affected the frequency of “Expected” responses (see Figure 2.4). There was a main effect of triad type $F(5, 70) = 18.55$, $p < .001$, $\eta p^2 = 0.57$. After Bonferroni corrections, pairwise comparisons revealed that triad vi was significantly more expected than all other triad types, and vii^{o} was less expected than ii, IV and V (see Table 2.5 for *p*-values for pairwise comparisons, corrected by Bonferroni corrections). There were no main effects or interactions associated with musical training (*p*s > .05).

Table 2.5 – *p*-values (after Bonferroni corrections) for pairwise comparisons between triad types for the related harmony chord type (a "–" indicates that the *p*-value was > .05).

Triad	ii	iii	IV	V	vi
ii					
iii	–				
IV	–	–			
V	–	–	–		
vi	<.001	.020	.027	<.001	
vii°	–	.010	<.001	.002	<.001

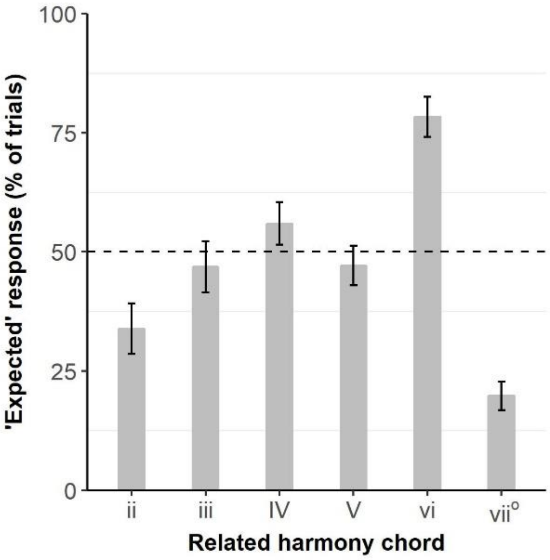


Figure 2.4 – Mean % of trials with an "Expected" response, for the six related harmony triad types: ii, iii, IV, V, vi and vii°. The dashed line at 50% illustrates the 50% (chance) level. Within-subject *SE* bars are included.

Reaction time

A third, 3 x 2, mixed ANOVA was carried out to investigate whether chord progression end type (3 levels within-subject: tonic (I), related harmony and unrelated harmony) or musical training (2 levels between-subject: amateur musicians and non-musicians) affected reaction time (see Figure 2.5). Mauchly's test indicated that sphericity had been violated χ^2 (2) = 9.85, $p = .007$. Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates ($\varepsilon = .65$). There was a significant effect of chord progression end type on reaction time $F(1.31, 18.29) = 22.77$, $p < .001$, $\eta p^2 = 0.62$. After Bonferroni corrections, pairwise comparisons revealed that reaction time was faster for both the tonic (I) chords *(M* = 503 ms, *SD* = 58 ms, $p < .001$) and unrelated harmony chords (568 ms, *SD* = 63 ms, $p = .001$) compared to the related harmony chords *(M* = 814 ms, *SD* = 197 ms). This effect could reflect task difficulty as the decision of whether the final chords were "Expected" or "Unexpected" was more ambiguous for related harmony chords, compared to the other two chord types. There were no main effects or interactions associated with musical training (*p*s > .05).

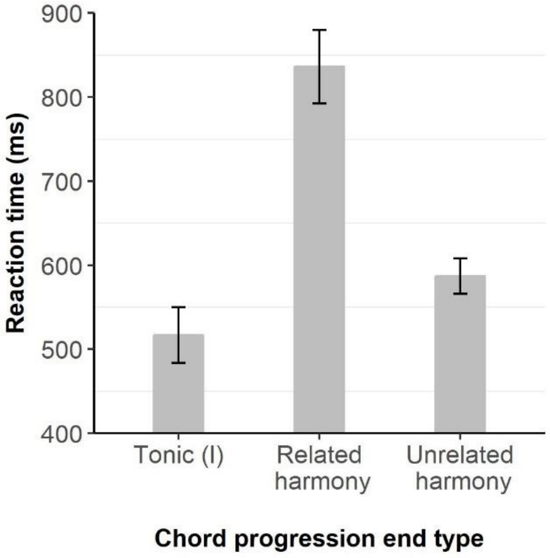


Figure 2.5 – Mean reaction time (ms) for the three chord progression end types: tonic (I), related harmony and unrelated harmony. Within-subject *SE* bars are included.

2.2.4. Experiment 1: Discussion

The main aim of Experiment 1 was to identify suitable stimuli to use as harmonic expectancy violations in the rest of the experiments presented in this thesis. The results of Experiment 1 confirm that the method used to create harmonic expectancy violations was successful. Participants responded mostly “Expected” to tonic (I) chords, mostly “Unexpected” to unrelated harmony chords and had a mixture of “Expected” and “Unexpected” responses to related harmony chords. Therefore, a successful method for creating harmonic expectancy violations was identified: using the tonic (I) for expected chords, a different triad in the key (ii, iii, IV, V, vi, vii °) for the related harmony chords and moving the related harmony chords up or down one semitone to create the unrelated harmony chords.

The results also shed light on differences in expectedness for different triads within the key (ii, iii, IV, V, vi, vii °), which has not been systematically investigated and reported in previous literature (Guo & Koelsch, 2016; James et al., 2008; Janata, 1995; Kim et al., 2011; Poulin-Charronnat et al., 2006; Steinbeis et al., 2006). The qualitative order of expectancy of the related harmony chords from most expected to least expected was: vi, IV, V, iii, ii, vii °. As predicted, this order is consistent with the count of triads following a dominant (V) chord in Bach’s chorales (order = vi, IV, iii, ii vii °[6]) (Rohrmeier & Cross, 2008). This result supports two hypotheses: first, Bach’s chorales provide a good approximation of harmony across Western tonal music, as previously hypothesised by Trainor and Trehub (1994) and second, that expectations of harmony are formed implicitly through music listening.

Moreover, the results demonstrate a fit between participants’ responses and harmonic theory, as the submediant triad (vi) was rated as significantly more “Expected” than all other triad types and the leading tone (vii °) was rated as significantly less “Expected” than most other triads (ii, IV and V). This fits with harmonic theory as the

---

[6] Note that Rohrmeier and Cross (2008) were interested in chord changes and therefore did not include a dominant (V) chord following a dominant (V) chord as it was not relevant for the purpose of their analysis.

submediant (vi) is the relative minor triad, so is the most closely harmonically related triad to the tonic (I). On the other hand, the leading tone (vii °) is the least closely harmonically related triad to the tonic (I). Therefore, the submediant triad (vi) could be suitable for future studies investigating subtle harmonic expectancy violations and the leading tone (vii °) could be suitable for studies investigating more severe harmonic expectancy violations that are still in the original key.

To identify suitable stimuli for the later experiments in this thesis, the aim was to identify related harmony chords that elicited an even mixture of “Expected” and “Unexpected” responses (with the frequencies of “Expected” responses nearest to 50%). This was the aim because related harmony chords were intended to investigate ERP effects for harmonic expectancy violations that are not severe in the later experiments. Experiment 1 showed that the triads with the most balanced (nearest to 50%) set of “Expected” and “Unexpected” responses were the mediant (iii), subdominant (IV) and dominant (V). As the dominant (V) is a repetition of the previous chord (all chord progressions originally ended on a perfect (V–I) cadence, and only the final chord (I) was changed), triads iii and IV are used for the related harmony chords in the subsequent experiments presented in this thesis.

Crucially, there were no differences in expectancy scores between musicians and non-musicians, for either chord progression end type (tonic (I), related harmony and unrelated harmony) or chord type (ii, iii, IV, V, vi, vii °). There were also no group differences in reaction time. These results reinforce that harmonic expectation is implicitly learned through listening to Western music and does not require formal musical training or expertise (Bigand & Poulin-Charronnat, 2006; Jonaitis & Saffran, 2009). As both musicians and non-musicians rated the expectancies of the chord progressions in the same way, these stimuli can be used to meaningfully compare neural responses between musicians and non-musicians in future studies. Any ERP differences between musicians and non-musicians in future studies are therefore unlikely to be confounded by differences in perceived expectancy of the stimuli.

2.3. Conclusion

To sum up, the expectancy violation stimuli created and tested in this chapter are similar in length, number, and presentation timing between language and music domains. This facilitates attempts to present analogous task paradigms and achieve a comparison of ERP expectancy effects, which could provide original insights into the age-old debate about shared and separable neurocognitive substrates between language and music domains. These stimuli are used to explore ERP correlates of semantic, grammar and harmonic expectancy violations in Experiments 2, 3, 6 and 7. They are adapted in Experiment 4 to test the expectancy of different levels of meter violations. From these adaptations, suitable stimuli are selected to test ERP correlates of meter expectancy violations in Experiments 5, 6 and 7.

3. Chapter 3 – ERP correlates of semantic, grammar and harmonic expectancy violations

3.1. Introduction

In the previous chapter, semantic, grammar and harmonic expectancy violations were systematically created. In this chapter, ERP effects associated with these violations are investigated. If the experiments provide evidence that ERP effects for semantic, grammar and harmonic expectancy violations are distinct, then the hypothesis that there are separable neurocognitive substrates involved in these processes could be supported.

3.1.1. Aim 1: Is there evidence of distinct ERP effects?

The current literature, discussed in detail in Chapter 1, suggests that the neurocognitive substrates underlying semantic processing might be domain-specialised and that a robust ERP marker of semantic expectancy violations is the N400 effect (Besson et al., 1998; Gutierrez et al., 2012; Kamp et al., 2015; Kutas & Federmeier, 2010; Kutas & Hillyard, 1983; Moreno & Vázquez, 2011; Nigam et al., 1992; Rommers et al., 2013; Tiedt et al., 2020). On the other hand, ERP effects associated with grammar and harmony processing, particularly the P600 and P300 effects, might indicate domain-general neurocognitive substrates (Patel, 2008; Patel et al., 1998; Sassenhagen & Fiebach, 2019). This hypothesis is logical because grammar and harmonic violations are violations of structure, and it follows that there might be domain-general neurocognitive substrates involved in structural processing.

Experiments 2 and 3 aimed to investigate whether the semantic, grammar and harmonic expectancy violation stimuli, created in Chapter 2, elicit distinct ERP effects and, therefore, indicate domain-specialised neurocognitive substrates in language and music domains. These stimuli were based on previous experiments that elicit N400 effects for semantic violations (Kutas & Hillyard, 1983), P600 effects for grammar violations (Gouvea et al., 2010), and P300 effects for harmonic violations (Janata, 1995). Therefore, it was predicted that these ERP effects would occur in the current experiments.

Several previous studies have attempted to find an N400 for harmonic expectancy violations, but none have found one (Besson & Faïta, 1995; Besson & Macar, 1987; Miranda & Ullman, 2007; Paller et al., 1992). Furthermore, N400s are not elicited when grammar expectancy violations are more prominent than semantic expectancy violations (Kim & Osterhout, 2005). Therefore, for the current experiments, it was predicted that the N400 would only occur for semantic expectancy violations.

It is less clear what to predict for the ERP effects typically elicited by grammar (P600) and harmonic (P300) violations. Notably, it has been suggested that the P600, found for grammar expectancy violations, might indicate the same neurocognitive substrates as P300 effects in the language domain (Sassenhagen & Fiebach, 2019; Sassenhagen et al., 2014). Despite these findings and the fierce debate about the potential overlap of mechanisms involved in grammar and harmony processing, there appear to be no previous active tests to compare P600 effects for grammar processing and P300 effects for harmony processing. Furthermore, although most studies do not report a P600 effect for harmonic expectancy violations, the classic study by Patel et al. (1998) reports one. The current experiments aimed to test whether there is a P600 effect for harmonic expectancy violations, and aim to provide insights into whether there are comparable scalp topographies (that could, in turn, indicate shared neurocognitive substrates) for ERP effects associated with grammar and harmonic expectancy violations.

Comparing ERP results across studies is not sufficient to test whether they are process-specific, as differences between them might be due to differences in samples, acquisition methods, task design or analysis protocols. With this in mind, in Experiments 2 and 3, a within-subjects design was adopted, with systematic stimuli design and an attempt at analogous task paradigms between language and music tasks. However, there was another methodological issue to overcome. When words are presented in the auditory modality, their onset lacks temporal precision they take some time to complete. This contrasts to chords in the music domain, which, immediately convey harmonic information when all notes in the chord are played simultaneously. The onset of language stimuli is easier to control in the visual modality, compared to the auditory modality, as whole words can be presented at once (Chee et al., 1999). Therefore, presenting language stimuli in the visual modality could aid comparisons of ERP effects for language and music stimuli. This approach is supported, as previous research has found that the N400 effect for semantic violations and P600 effect for grammar expectancy violations are identical when presented in visual and auditory modalities, suggesting that they are modality-independent (Balconi & Pozzoli, 2005; Kutas & Federmeier, 2000). With these findings in mind, the language stimuli are presented visually in the current experiments.

Furthermore, classic statistical techniques for analysing ERP effects are limited in the information they provide about temporal and spatial characteristics. They focus on a-priori selected single electrodes or groups of electrodes and specific time windows to avoid the problem of multiple comparisons (Luck, 2014). With new techniques, such as cluster-based permutation statistics (Fields & Kuperberg, 2020), it is possible to test all scalp electrodes and whole epochs for significant differences between conditions in one test, minimising the multiple comparisons problem. This method allows for ERP effects that are not predicted a-priori to be identified and could provide novel insights into shared ERP effects for different types of expectancy violations. Lastly, with a within-subjects design, it is possible to actively test whether any effects found for one type of expectancy violation also occur for other types of expectancy violations. Together, these methods enable original conclusions about the process-specificity of ERP effects for expectancy violations in language and music domains which could, in turn, give insights into domain-specialised or domain-general neurocognitive substrates in language and music domains.

3.1.2. Aim 2: Comparing task difficulty between language and music tasks

Previous studies suggest that differences in task difficulty could limit the ability to draw conclusions about differences in ERP effects (Magne et al. 2005). With this in mind, in the current experiments, task difficulty will be compared between language and music tasks, with longer reaction times taken to suggest greater task difficulty (Palmer et al., 1994; Warrick et al., 1965).

3.1.3. Aim 3: Assessing task paradigms

In both language and music domains, it remains unclear how expectancy violation processing is affected by task paradigm choice. The tasks presented in the literature broadly fall into two categories: explicit and implicit. In explicit tasks, detection of the expectancy violation is task-relevant. In language tasks, for example, participants are required to decide, with a “Yes” or “No” answer, whether sentences are “Correct” or “Incorrect” (Hahne et al., 2012; Li et al., 2018) or to rate the congruence between a visual scene and a sentence on a Likert scale (Coco et al., 2017). Common explicit music tasks require participants to respond with “Yes” or “No” to the question whether the chord progression is “Beautiful” (Müller et al., 2010) or “Satisfactory“ (James et al., 2008). On the other hand, in implicit tasks, the expectancy violation is task-irrelevant. For example, participants can be simply asked to read sentences, without requiring participants to overtly respond to the expectancy violations themselves (Relander et al., 2009). Alternatively, participants might be told they would answer questions about the sentences at the end of the experiment (Goregliad Fjaellingsdal et al., 2016; Kamp et al., 2015; Kutas & Hillyard, 1983). Indeed, this method is often adopted in attempt to encourage the participants to pay attention to the words. For music experiments, implicit tasks often involve focusing on a silent movie (rather than the musical stimuli) and, in some cases, participants are asked to try to ignore the auditory music violations (Krohn et al., 2007).

Previous studies have suggested that the task relevance of the expectancy violation can have notable effects on their neural correlates. For example, Erlbeck et al. (2014) conducted a study in which participants were either asked to focus their attention on the sentences, passively listen to the sentences or ignore the sentences. The N400 effect was smaller for the passive task compared to the focused task and was not elicited in the ignore task. Similarly, the P600 tends to be larger when the grammar violations are task-relevant, and it tends to become small or not to occur when they are task-irrelevant (Hahne & Friederici, 2002; Haupt et al., 2008; Lemhofer et al., 2020; Osterhout et al., 2002).

ERPs elicited by harmonic violations can also be affected by choice of task paradigm. P300 effects have been shown to be more likely to occur when the task makes expectancy violations explicit (Friedman et al., 2001). Furthermore, Ellison et al. (2015) included two types of tasks in their experiment: participants listened to chord sequences that were either harmonically correct or harmonically incorrect. They reported a late positive ERP potential when the task was to respond to whether chord sequences were "Beautiful" or "Not beautiful", but a late negative ERP potential was elicited when the task was to respond "Correct" or "Not correct". It is, therefore, necessary to be aware of the potential impacts of the chosen task paradigm to be able to appropriately attribute ERP effects to the experimental manipulation at hand.

In the current chapter, different task paradigms were designed for Experiment 2 and Experiment 3. Experiment 2 adopted a task paradigm where the expectancy violations were task-relevant: participants were asked whether the final (target) word or chord were "Expected?" or "Unexpected?", to which they responded "Yes" or "No". In contrast, Experiment 3 adopted task paradigms where the expectancy violations were task-irrelevant. For the language task, participants were asked whether the sentence was "Green?" or "Blue?" and they answered "Yes" or "No", and for the music task, participants were asked whether the chord progression was played by the "Piano?" or "Organ?" and they answered "Yes" or "No". Including both tasks reveals first whether the ERP effects are robust and replicable, and second whether they are affected by the task-driven saliency of the expectancy violations.

3.1.4. Aim 4: Testing the expectancy of the language violation stimuli

The chord progression endings identified based on how "Expected", or "Unexpected" participants rated them in Experiment 1 were used as the music stimuli in the current experiments. It was necessary to complete Experiment 1 due to limited previous research on creating expectancy violations in the music domain. The language stimuli were systematically created based on a rich body of prior research on creating semantic and grammar expectancy violations (described in section 2.1.1). As a result, prior to Experiment 2, it was not deemed critical to test how “Expected”, or “Unexpected” people find these stimuli. However, understanding the perceived expectancy of expectancy violation stimuli aids the interpretation of associated ERP effects. Therefore, Experiment 2 aimed to test how “Expected” and “Unexpected” participants rated these sentence endings.

3.1.5. Aim 5: Investigating expertise effects

There is evidence of within-domain expertise effects on ERP effects for expectancy violations. For example, there have been reports of larger N400s for semantic expectancy violations in native speakers compared to non-native speakers (Anurova & Immonen, 2017), larger P600 effects for grammar violations for people who were more proficient in a novel artificial language (Batterink & Neville, 2013), and larger P300 effects for harmonic violations for musicians compared to nonmusicians (George & Coch, 2011; Steinbeis et al., 2006). Crucially, limited previous research also demonstrates evidence of cross-domain expertise effects. For example, Dittinger et al. (2017) found that in children with musical training, the N400 to unexpected newly-learned words was significantly larger than in children without musical training. In the current experiments, analyses were conducted to test for within- and cross-domain differences between native and non-native English speakers and between amateur musicians and non-musicians in any significant effects of interest (expectancy violation ERP effects). Cross-domain expertise effects could indicate shared neurocognitive substrates between language and music domains (Patel, 2011).

3.2. Experiment 2: ERP correlates of semantic, grammar and harmonic expectancy violations – an implicit task

3.2.1. Experiment 2: Method

Ethics

The University Teaching and Research Ethics Committee at The University of St Andrews approved this study, code PS13256 (see Appendix P).

Participants

Twenty university students volunteered to participate in this study (4 male, Age: *M* = 21 years, *SD* = 2 years, range = 17–27 years). Descriptive statistics of participants' English language and musical training experience are in Table 3.1. Each testing session took 120 minutes, and participants were reimbursed £10. People could not sign up to participate in the study if they had participated in Experiment 1.

Table 3.1 – Descriptive statistics of participants' English language and musical training experience, Experiment 2.

Experience	Measure	*M* (*SD*)	Range
English language	Native English speaking	Native: 12, Non-native: 8	
	Onset (age in years)	2.4 (3.4)	0-10
	English ability (rating 1–10)	9.4 (1.0)	7-10
	% English spoken in the past year	82.2 (21.8)	45-100
Musical training	Number of years practising	7.0 (5.1)	1–14
	Practice time (hours)	1695 (2218)	0–8424
	Overall ability rating (1–5)	3.1 (1.3)	1–5
	Listening per week (hours)	22.9 (18.6)	3–70

Apparatus

Stimuli presentation and data collection was run on Experimental Run Time System Version 3.32 (Beringer, 1994). The stimuli were presented at a viewing distance of roughly 80 cm on a 17–inch cathode ray tube (CRT) monitor. Responses were recorded using a keypad with two keys set 10 cm horizontally apart. The EEG system was a Biosemi Active-Two amplifier system with 72 Ag/AgCl electrodes. A common mode sense (CMS) active electrode was used as a reference electrode and a drive right leg (DRL) passive electrode as a ground electrode. Electrolyte gel was used to improve conductivity between the scalp and the electrodes. EEG was recorded at a sampling rate of 256 Hz. There were 64 scalp electrodes, four electrodes recorded electro-oculographic activity, and two electrodes were placed over the right and left mastoids (see Figure 3.1).

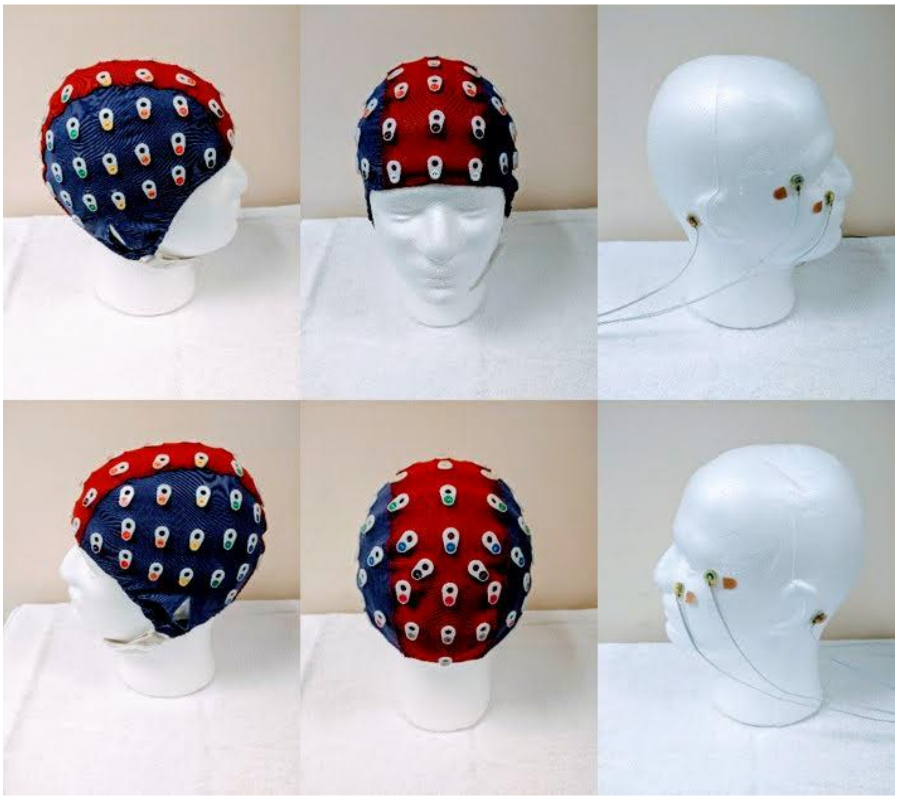


Figure 3.1 – The EEG cap and external electrodes set-up.

Materials

Sentences

In the language task, 192 sentences were presented. As designed in section 2.1.1, sentence stems were adapted from those presented by Block and Baldwin (2010), who investigated the cloze probability of 498 sentences (*N* = 400). The 24 sentences stems were presented with four sentence end types: standard words, related semantic violation, unrelated semantic violation and grammar violation. Endings with high cloze probability (> 90%), as identified by Block and Baldwin (2010) were used for the "standard word" endings. For the related semantic violation manipulation, words that were technically semantically correct but not likely were chosen. For the unrelated semantic violation manipulation, semantically incorrect words were used. For the grammar expectancy violation, a word with a similar semantic meaning to the standard word, but with incorrect grammar was used. Table 3.2 shows an example of the four sentence end types (see Appendix A for the full list of sentences):

Table 3.2 – An example of a sentence stem with the four sentence end types

Sentence stem example	Sentence ending	Sentence end type
He washed the dirty dishes in the	sink.	Standard word
	bathtub.	Related semantic
	pencil.	Unrelated semantic
	sinking.	Grammar violation

The 96 (24x4) sentences were each presented twice, one time to ask "Expected?" and the other time to ask "Unexpected?". There were also 12 practice trials, for which three sentence stems that were not used for the main experiment were repeated with the four sentence end types. In both the practice trials and the main experimental trials, half of the practice trials showed the question "Expected?" and half showed "Unexpected?", to which participants answered "Yes" or "No".

Chord progressions

In the music task, 144 chord progressions were presented. Twenty-four chord progression stems were used from Experiment 1, which were developed from passages in Bach's chorales (Bach-Gesellschaft, 1892). The 24 chord progression stems were presented with three chord progression end types: tonic (I), related harmony and unrelated harmony. The related harmony chords consisted of only mediant (iii) and subdominant (IV) chords, as they were identified as the most appropriate chords in Experiment 1. These related harmony chords are within the key of the rest of the chord progression but are not the anticipated (tonic, I) ending. The unrelated harmony chords corresponded to the related harmony chords – the related harmony chords were either moved up or down one semitone. This method created a chord that did not fit within the harmonic context of the chord progression. 12 of the 24 chord progressions were played by a piano sound and 12 by an organ sound. For both sets of 12, there was one chord progression in each major key signature. Seven chord progressions were eight chords in length, and five were seven chords in length. The 72 chord progressions (24x3) were each presented twice, one time to ask “Expected?”, and the other time to ask “Unexpected?”. In addition, there were 12 practice trials, for which four chord progression stems that were not used for the main experiment were repeated with the three chord progression end types. As for the main experiment, half of the practice trials asked “Expected?” and half asked “Unexpected?”. Participants responded “Yes” or “No”. For an illustration of the chord progressions, with the three chord progression end types, see Figure 2.1.

Additional materials

English Language Experience Questionnaire: A self-report questionnaire was designed to measure proficiency in the English language and other spoken languages (see Appendix D). It requires participants to record how many (and what) languages they speak fluently, their ability in each language on a scale from 1 = not proficient to 10 = native-like ability, the age they started learning each language, and what percentage of time they have used each language in the past year.

Musical Training Questionnaire: See section 2.2.2 Experiment 1: Method (Materials).

Procedure

After reading an information sheet and signing a consent form, participants were prepared for EEG recording. Participants sat in an electrically shielded booth. Auditory material was presented over loudspeakers, and visual material was presented on a computer screen. There were two tasks: the language task and the music task. The order of the tasks was counterbalanced between participants. In the language task, there were 12 practice trials and 192 experimental trials. The sentences were presented in random order. The language task was presented in eight blocks of 24 trials each. After the language task, participants were asked to answer five questions about the sentences to check that they paid attention to the sentences. In the music task, there were 12 practice trials and 144 experimental trials. The chord progressions were presented in random order. The music task was presented in six blocks of 24 trials each. Participants were able to take breaks after each block.

For each sentence or chord progression, one word or chord was presented every 1000 ms (see Figure 3.2). 1000 ms after the final word or chord of each sentence or chord progression, participants were asked whether the final word or chord was “Expected?” or “Unexpected?” to which they responded “Yes” or “No” using the response keys. Responses were only accepted for analysis if they were given between 100 ms and 5000 ms after question onset.

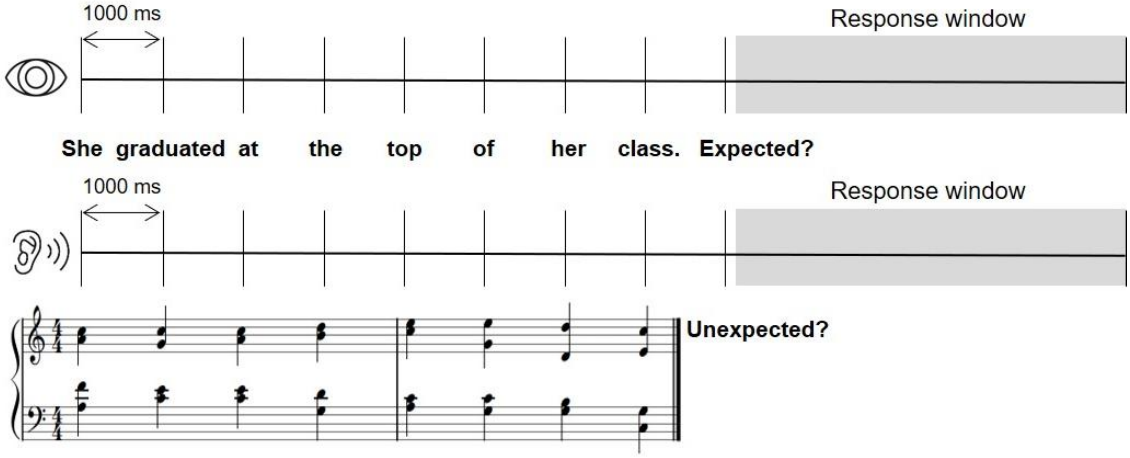


Figure 3.2 – The timing of stimuli presentation for the language task (top) and the music task (bottom). The examples use the standard (expected) word, and tonic (I) chord endings. The response window (100–5000 ms after question onset) is shaded in grey.

Participants chose whether they answered “Yes” with the right hand and “No” with the left (*N* = 12), or vice versa (*N* = 8). This response paradigm was chosen (instead of assigning, e.g., “Expected” left-hand response and “Unexpected” right-hand response) so that participants were unable to anticipate which motor response they would use until they saw the question word. Therefore, there should be no potentially confounding effects of anticipation of motor response in the EEG signal. Additionally, the trials were presented in randomised order so that participants could not prepare a specific response before the full sentence or chord progression had been presented. After both tasks were completed, participants filled in the English Language Experience Questionnaire and the Musical Training Questionnaire. It was deemed important that participants fill out the questionnaires after the task was completed, in case increasing the saliency of their English language experience or musical training experience (or lack thereof) affected their responses.

Analysis

For details of the analysis and graphing software and statistical correction procedures used, see section 2.2.2 Experiment 1: Method (Analysis). ERPLAB functions were used to plot the ERP waveform and scalp topography maps (Lopez-Calderon & Luck, 2014).

EEG preprocessing

EEG data preprocessing was done in EEGLAB and ERPLAB (Delorme & Makeig, 2004; Lopez-Calderon & Luck, 2014). The epochs were extracted from the continuous EEG recording, starting 200 ms before and ending 1000 ms after the onset of the final word or chord. The data was then high-pass filtered at 0.1Hz. Bad channels (channels with poor signal quality, usually due to a poor-quality connection between the electrode and the scalp) were removed from the data, and a temporary Cz reference applied. Artefact correction was done using independent components analysis (ICA) and the ADJUST function in ERPLAB (Mognon et al., 2011). All remaining artefacts were removed using the moving peak-to-peak window method in ERPLAB. The epochs were then baseline corrected (–200 to 0 ms relative to the onset of final word or chord) and re-referenced to average reference. Epochs were averaged separately for each experimental condition and participant, and a low-pass filter of 30Hz applied. For more information about preprocessing, and for justification of these choices, see Appendix E.

Trial inclusion

Only trials with responses given from 100–5000 ms after the onset of the question prompt were considered for analysis. As there was no strictly correct or incorrect response to the task, all trials that were responded to within the time frame were considered for analysis. Table 3.3 shows the mean and standard deviation of how many trials were left in the analysis after the behaviour and artefact rejection inclusion criteria. Luck (2014) suggested that ERP studies should be designed with between 30–40 trials per condition with 20 participants to have an acceptable signal-to-noise ratio to detect late-onset ERP components, assuming that between 10–25% of these will be rejected due to artefacts. In Experiment 2, all end types had more than 42 trials, with 20 participants, and, therefore, the dataset was deemed to have an acceptable signal-to-noise ratio for analysis.

Table 3.3 – Mean (*SD*) number of trials included in the analysis after the inclusion criteria were applied in Experiment 2. The maximum possible number of trials for each end type was 48.

		Mean number of trials (*SD*)	
Task	End type	Behaviour	Artefact rejection
Music	Tonic (I)	47.9 (0.3)	43.5 (3.8)
	Related harmony	47.4 (1.4)	43.5 (4.3)
	Unrelated harmony	47.7 (0.6)	42.6 (4.3)
Language	Standard word	47.8 (0.4)	43.2 (4.6)
	Related semantic	47.9 (0.5)	44.2 (4.3)
	Unrelated semantic	47.5 (0.7)	43.5 (4.1)
	Grammar violation	47.9 (0.5)	43.9 (4.4)

Analysis of ERP effects

The ERP effects were analysed with two methods: cluster-based permutation statistics and ANOVAs. Cluster-based permutation statistics are a relatively new method in ERP analysis (Fields & Kuperberg, 2020). This method is beneficial because it avoids the multiple comparisons problem and it has the ability to identify multiple effects within a large epoch in one test. This method provides weak control of the family-wise error rate (FWER) and is generally best at identifying less focal, later-onset effects. The “mass” of a cluster is the sum of the *F*-values in that cluster. Factorial Mass Univariate ERP Toolbox (Fields, 2017), an extension of the Mass Univariate ERP Toolbox (Groppe et al., 2011), was used to calculate the cluster-based permutation statistics. These toolboxes support within-subjects analyses. Clusters were created when *F*-values were significant in spatiotemporally adjacent samples. The maximum distance between electrodes was 3.3 cm, as this is the maximum distance between electrodes on a 64–Biosemi cap, assuming standard head size with a diameter of 56 cm. For each test, there were 10,000 random permutations. The threshold for the *p*-value was .05. Therefore, all clusters with *p*-values < .05 were defined as statistically significant. All cluster-based permutation tests were focused between 0–800 ms after the target stimulus onset.

ANOVAs are the classic method for ERP analysis and are usually based on a-priori selected electrodes and time windows (Luck, 2014). In the current experiments, suitable electrodes for ANOVA analysis were identified with the cluster results, chosen near the spatial peak of significant clusters. This was deemed to be the most suitable method because the methods for choosing key electrodes are not consistent across previous studies, and the scalp topographies of these effects are not always well-defined. This is particularly the case for the P300 effect, as there are thought to be different types of P300 effect associated with harmonic violations of different types (the P3a and P3b), and distinctions between their scalp topographies have not been clearly defined (Polich, 2007).

In contrast, time windows for the ERP ANOVA analyses were chosen a-priori, as these tend to be more consistent across studies in the previous literature. A-priori time windows were chosen for the predicted effects: the N400 effect for semantic violations, P600 effect for grammar violations and P300 effect for harmonic violations. Most N400 a-priori time windows fall somewhere between 250–500 ms, but they are often shorter in an attempt to increase temporal information about the effect (Kutas & Federmeier, 2010). With this in mind, for the current experiment, the N400 a-priori time window was 300–450 ms, 50 ms shorter than the average on both sides. The P600 effect usually begins around 500 ms after the onset of a grammar expectancy violation, reaches its peak around 600 ms and continues for several hundreds of milliseconds (Gouvea et al., 2010). For this reason, in the current study, the a-priori P600 time window was set at 500–800 ms (800 ms was the end of the tested epoch). The classic P300 peak is often measured between 250–400 ms (Janata, 1995). This is the a-priori time window used to test for the P300 effect in the current study. For the ANOVAs, the mean amplitudes in the specified time ranges were calculated. It is most appropriate to use the mean amplitude method for these later components, rather than, e.g., peak amplitude or peak latency, because they do not always have clear and well-defined peaks. The a-priori time windows for analysis are shown in Table 3.4.

Table 3.4 – A-priori chosen time windows for analysis of N400, P600 and P300 components.

ERP component	Time (ms)	Violation type
N400	300–450	Semantic
P600	500–800	Grammar
P300	250–400	Harmonic

The cluster-based permutation statistics were used to identify significant clusters, and from the results of these clusters, suitable electrodes were chosen for ANOVA analysis. This method was chosen because scalp topographies of expectancy violation ERP effects tend not to be well-defined in the literature, and an aim of the current experiment was to provide information about the scalp topography of these ERP effects. It was hoped that, in turn, this might shed light on the distinctness of these ERP effects, and support inferences about the distinctness of their associated neurocognitive substrates.

Nieuwenhuis et al. (2011) suggest that to argue that two ERP components are distinct for different stimuli, it is necessary to establish that for each effect, at a key electrode and in the time window of analysis, there is no similar effect for the stimuli that elicited the other component. In the current experiments, for example, it needs to be shown that for any ERP effects of interest for semantic expectancy violations, there is no comparable ERP effect for grammar or harmonic expectancy violations. To test for this, further ANOVA tests were carried out, using the selected electrodes and a-priori time windows.

Analysis of expertise effects

The analyses first include all participants to retain statistical power to detect ERP effects associated with expectancy violations. For any significant expectancy violation, ERP effects, a between-subjects factor of English language experience (2 groups: native and non-native speakers) and a between-subjects factor of musical training (2 groups: amateur musicians and non-musicians) were added to the ANOVA designs to test expertise effects. In the studies presented in this thesis, to explore language expertise effects, native English speakers were compared with non-native English speakers, as previous studies have found differences in language processing between native and non-native speakers Anurova and Immonen (2017); (Cuskley et al., 2015; Kisser et al., 2012; Trenkic et al., 2014). Musical expertise was measured, with a cut-off of > 1000 hours of accumulated practice time for amateur musicians and < 1000 hours of accumulated practice time for non-musicians. For the rationale of this cut-off point, see Appendix F.

3.2.2. Experiment 2: Behaviour results

Response Frequency

Language task

A within-subjects ANOVA tested whether the frequency of “Expected” responses was affected by sentence end type (4 levels: standard, related semantic violation, unrelated semantic violation and grammar expectancy violation) (see Figure 3.3, left panel). For Mauchly’s test, $\chi^2(5) = 34.66$, $p < .001$. Degrees of freedom were corrected using Greenhouse-Geisser estimates ($\varepsilon = .61$). Sentence end type affected how expected the final words were $F(1.84, 35.02) = 567.55$, $p < .001$, $\eta p^2 = 0.97$. The standard words received more “Expected” responses (M = 97%, SD = 11%) than the other sentence end types (related semantic violations M = 13%, SD = 9%, unrelated semantic violations M = 5%, SD = 3%, and grammar expectancy violations M = 8%, SD = 9%, all *p*s < .001). Additionally, the related semantic violations received more “Expected” responses than the unrelated semantic violations ($p = .005$). All other pairwise tests were not significant ($p > .05$).

Music task

A second within-subjects ANOVA tested the effect of chord progression end type (3 levels: tonic, related harmony and unrelated harmony) on the frequency of “Expected” responses, $F(2, 38) = 106.35$, $p < .001$, $\eta p^2 = 0.85$ (see Figure 3.3, right panel). All pairwise comparisons were significant at the $p < .001$ level. As in Experiment 1, the tonic

(I) chords received mostly “Expected” responses *(M* = 94%, *SD* = 18%), the related harmony chords received a mixture of responses *(M* = 50%, *SD* = 15%) and the unrelated harmony chords received the least percentage of “Expected” responses *(M* = 15%, *SD* = 18%).

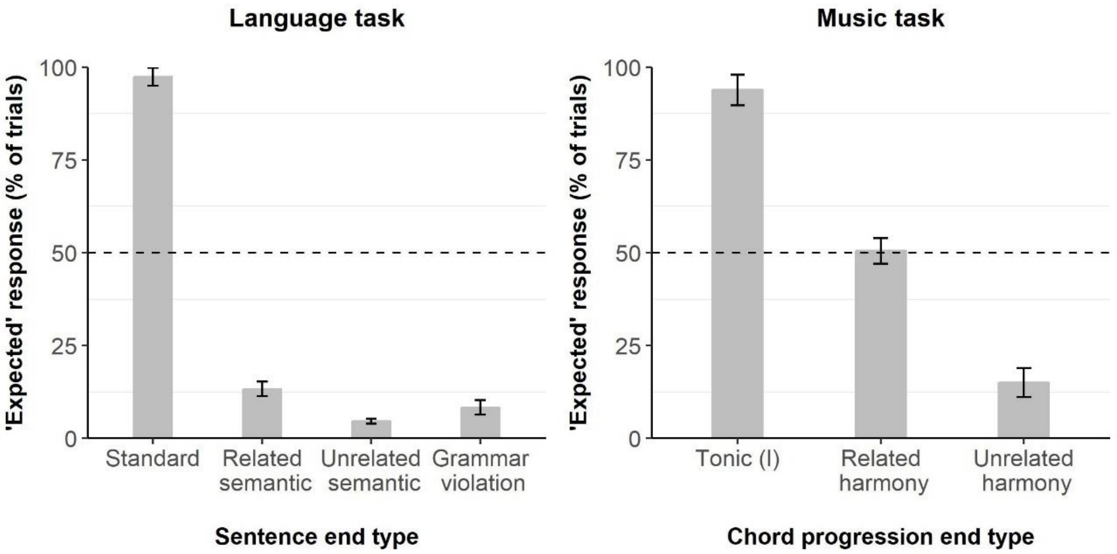


Figure 3.3 – Mean % of trials with an “Expected” response for each end type in the language task (left) and the music task (right). The dashed line illustrates the 50% (chance) level. Within-subject *SE* bars are included.

Reaction time

Language task

A within-subjects ANOVA tested whether reaction time was affected by the sentence end type (4 levels: standard, related semantic violation, unrelated semantic violation and grammar expectancy violation). There was an effect of sentence end type on reaction time *F*(3, 57) = 20.12, *p* < .001, ηp^2 = 0.51 (see Figure 3.4, left panel). The reaction time was significantly faster for the standard words *(M* = 966 ms, *SD* = 73 ms) than the related semantic violations *(M* = 1139 ms, *SD* = 81 ms, *p* <.001) and grammar expectancy violations *(M* = 1036 ms, *SD* = 75 ms, *p* = .026). The related semantic violations were also responded to slower than the grammar expectancy violations (*p* = .007) and unrelated semantic violations *(M* = 1006 ms, *SD* = 65 ms, *p* = <.001). These differences in reaction time appear to reflect task difficulty, as it was most ambiguous to judge the sentences as “Expected” or “Unexpected” for the related semantic violations, compared to the three other sentence end types.

Music task

A within-subjects ANOVA tested the effect of chord progression end type (3 levels: tonic, related harmony and unrelated harmony) on reaction time. It showed that chord progression end type affected reaction time $F(2, 38) = 31.54$, $p < .001$, $\eta p^2 = 0.62$. (see Figure 3.4, right panel). All pairwise comparisons were significant at the $p < .05$ level (see Table 3.5 for *p* values). Reaction time was fastest for the tonic (I) chord *(M* = 902 ms, *SD* = 90 ms), followed by the unrelated harmony chord *(M* = 976 ms, *SD* = 99 ms), and the slowest was for the related harmony chord *(M* = 1164 ms, *SD* = 130 ms). This reaction time effect could reflect task difficulty, as participants may have found it easiest to decide whether the tonic (I) was “Expected” or “Unexpected”, as they responded “Expected” on most of the trials *(M* = 94%, *SD* = 18%), and most difficult to decide about the related harmony chord, as this received a fairly even mixture of “Expected” and “Unexpected” responses *(M* = 50%, *SD* = 15%). There was also an expectancy effect as the tonic (I) chord, which was the least surprising event, was responded to more quickly than the unrelated harmony chord.

Table 3.5 – *p* values (after Bonferroni corrections) for pairwise comparisons of reaction time between the three chord progression end types: tonic (I), related harmony and unrelated harmony.

Chord progression end type	Tonic (I)	Related harmony
Tonic (I)		
Related harmony	<.001	
Unrelated harmony	.023	<.001

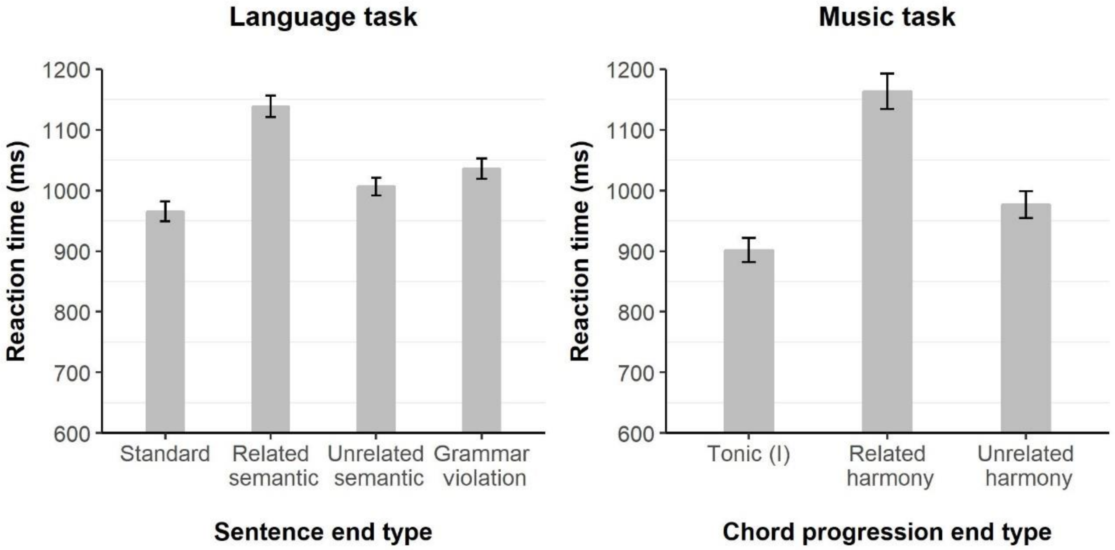


Figure 3.4 – Mean reaction time (ms) for each end type in the language task (left) and the music task (right). Within-subject SE bars are included.

Comparison between language task and music task

A *t*-test was conducted to compare participants' performance in the language and music tasks. Reaction time data were collapsed over end types in each task. There was no significant difference in reaction times between language and music tasks *t*(19) = 0.50, *p* = .631, *d* = 0.10. This result suggests that cognitive demand was similar between language and music tasks.

3.2.3. Experiment 2: ERP results

Language domain ERP results: Semantic expectancy violations

Within-subjects cluster-based permutation statistics with three levels of sentence end type (standard words, related semantic violations and unrelated semantic violations) identified a significant main effect cluster (mass = 799, *p* =.015), in centroparietal areas (C3, CP3, CP1, P1, P3, PO3, Pz, CP4, CP2, P2, P4) from 297–422 ms, see Figure 3.5. The temporal peak was 320 ms, and the spatial peak was at CP1. No additional clusters were identified. The midline Pz electrode was deemed suitable for ANOVA analysis. Sentence end type affected the mean amplitude at Pz in the N400 a-priori time window (300–450 ms), *F*(2, 38) = 6.12, *p* = .005, ηp^2 = 0.24. There was an increase in negativity for the unrelated semantic violations *(M* = 0.23 µV, *SD* = 0.32 µV) compared to the standard words *(M* = 1.22 µV, *SD* = 0.26 µV, *p* = .026). Pairwise comparisons revealed no main effects associated with related semantic violations (both *p*s > .05).

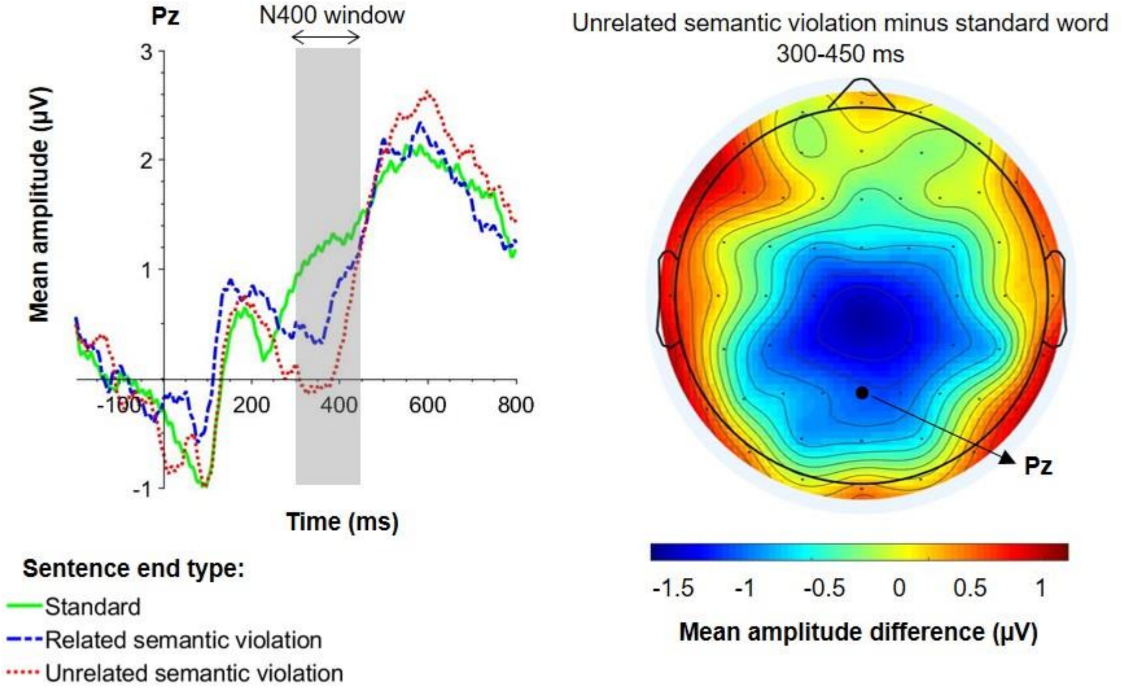


Figure 3.5 – N400 semantic expectancy violation graphs. Left: Mean amplitude (µV) at Pz between –200 to +800 ms for the three sentence end types. Right: Scalp topography map of the mean amplitude difference (µV) between 300–450 ms after the final (target) word's onset.

Language domain ERP results: Grammar expectancy violations

Within-subjects cluster-based permutation statistics with two levels of sentence end type (standard words and grammar expectancy violations) identified a significant cluster (mass = 1220, *p* =.046), in central areas (FC2, FC4, C2, C4, CP2, Cz, C1, CP1) from 398 ms until the end of the tested epoch (800 ms), see Figure 3.6. The temporal peak was 695 ms, and the spatial peak was at C4. No additional clusters were identified. The midline Cz electrode was deemed suitable for ANOVA analysis. A within-subjects ANOVA showed that sentence end type affected the mean amplitude at Cz in the P600 time window (500–800 ms), *F*(1,19) = 17.82, *p* = .001, ηp^2 = 0.48. The mean amplitude for the grammar expectancy violations was more positive *(M* = 1.82 µV, *SD* = 0.48 µV) than the mean amplitude for the standard words *(M* = 0.37 µV, *SD* = 0.34 µV).

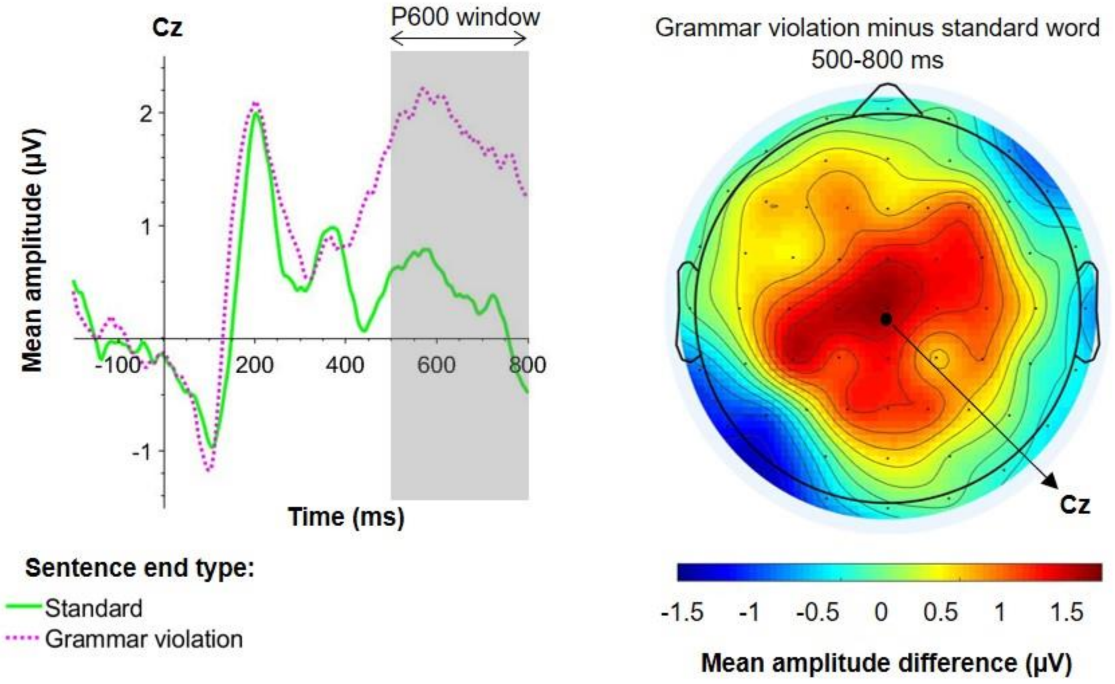


Figure 3.6 – P600 grammar expectancy violation graphs. Left: Mean amplitude (µV) at Cz between –200 to +800 ms for the two sentence end types: standard word and grammar expectancy violation. Right: Scalp topography map of the mean amplitude difference (µV) between 500–800 ms after the final (target) word’s onset.

Music domain ERP results: Harmonic expectancy violations

Within-subjects cluster-based permutation statistics with three levels of chord progression end type (tonic (I), related harmony and unrelated harmony chords) identified a cluster (mass = 2064, *p* =.013) with a frontal distribution (F1, F3, FC3, FC1, AF4, Fz, F2, F4, F6, FC4, FC2) which was significant from 211–461 ms (see Figure 3.7). The temporal peak was 352 ms, and the spatial peak was at FC2. The midline electrode Fz was chosen for ANOVA analysis, which showed that chord progression end type affected the mean amplitude at Fz in the P300 time window, between 250–400 ms *F*(2, 38) = 10.19, *p* <.001, $\eta p^2 = 0.35$. Mean amplitudes were more positive for both the related (*M* = 0.09 µV, *SD* = 0.22 µV, *p* = .004) and unrelated (*M* = 0.15 µV, *SD* = 0.27 µV, *p* = .002) chords compared to the tonic (I) chords (*M* = –0.98 µV, *SD* = 0.25 µV). There was no difference between the related and unrelated harmony chords (*p* = 1.00).

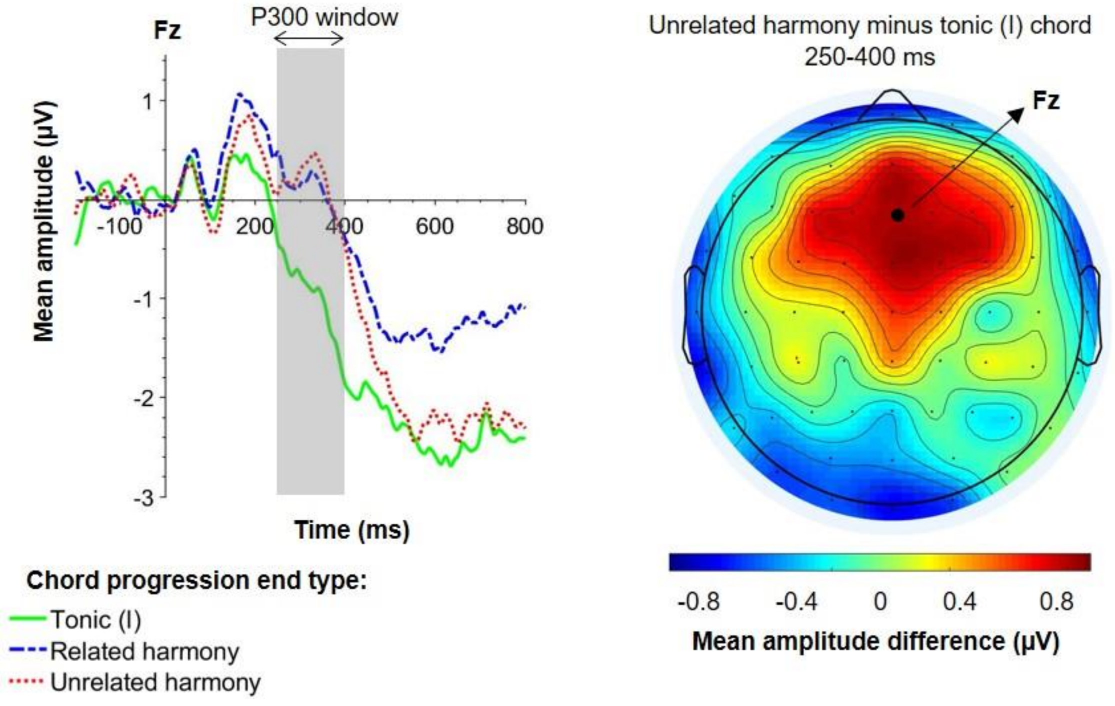


Figure 3.7 – P300 harmonic expectancy violation graphs. Left: Mean amplitude (µV) at Fz between –200 to +800 ms for the three chord progression end types. Right: Scalp topography map of the mean amplitude difference (µV) between 250–400 ms after the final (target) chord's onset.

There were two additional clusters (see Table 3.6 and Figure 3.8). Based on the characteristics of the clusters (Table 3.6), exploratory within-subjects ANOVAs with the three levels of chord progression end type: tonic (I), related harmony and unrelated harmony, were conducted at AFz between 200 and 800 and at POz between 220–800 ms. Measured at AFz, the frontal effect, *F*(2, 38) = 6.26, *p* = .004, ηp^2 = 0.25, was characterised by increased positive mean amplitude for the related harmony chords (*M* = –0.75 µV, *SD* = 0.48 µV) compared to the tonic (I) chords (*M* = –2.17 µV, *SD* = 0.50 µV, *p* = .032) and unrelated harmony chords (*M* = –1.63 µV, *SD* = 0.43 µV, *p* = .032). Measured at POz, the posterior effect, *F*(2, 38) = 16.54, *p* < .001, ηp^2 = 0.46, showed increased negative amplitude for the related harmony chords (*M* = 1.10 µV, *SD* = 0.36 µV) compared to the tonic (I) chords (*M* = 2.40 µV, *SD* = 0.42 µV, *p* = .001) and unrelated harmony chords (*M* = 2.59 µV, *SD* = 0.56 µV, *p* = .001). For both effects, there were no differences between tonic and unrelated chords (*ps* = .531 and .797, respectively)

Table 3.6 – Unforeseen cluster-based permutation statistics results for harmonic expectancy violations

Mass	*p*	Electrodes	Temporal Peak (ms)	Spatial peak	Extent (ms)
4773	<.001	T7, TP7, CP5, P1, P3, P5, P7, PO7, PO3, O1, Oz, POz, Pz, P2, P4, PO8, PO4, O2	547	PO3	Start: 203 End: –
2617	.001	Fp1, AF7, AF3, F1, F3, F7, FC3, FC1, Fpz, Fp2, AF8, AF4, AFz, Fz, F2, F4, F6, F8, FT8, FC6, FC4, FC2, T8, TP8	398	FC2	Start: 219 End: –

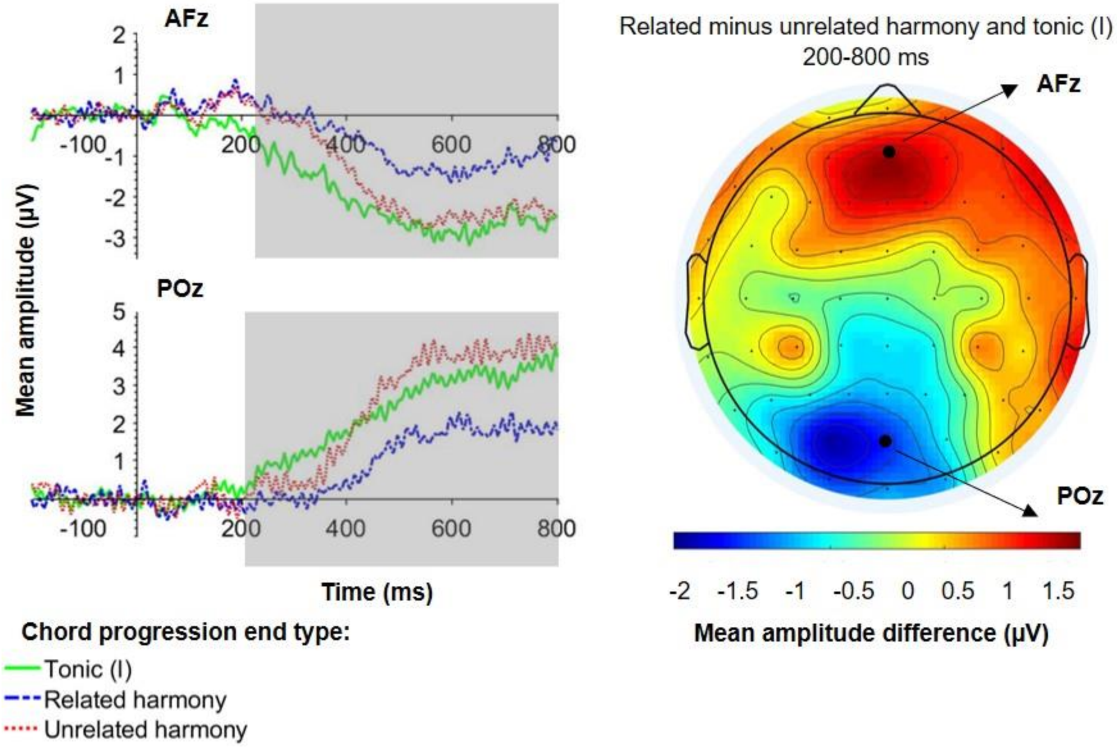


Figure 3.8 – Additional clusters in the music task graphs. Left: Mean Amplitude (µV) at Pz between –200 to +800 ms. Right: Scalp topography map with mean amplitude difference (µV) between 400–800 ms.

3.2.4. Experiment 2: Test of distinct ERP effects

The results discussed in the previous section show that there were clusters associated with an N400 effect for semantic expectancy violations, a P600 effect for grammar expectancy violations and a P300 effect for harmonic expectancy violations. Nieuwenhuis et al. (2011) suggest that, in order to strengthen an argument that two types of stimuli elicit different ERP effects that indicate different neurocognitive substrates, it is necessary to establish that for each effect, there is no comparable effect for the other stimulus type. Therefore, ANOVA analyses were run that tested each time window and electrode of interest that were identified in the current experiment with the other types of expectancy violations. The results of all these analyses were not significant (all *ps* > .05), see Appendix G for full details. Altogether, the evidence suggests that there are distinct ERP effects for semantic, grammar and harmonic expectancy violations. These distinct ERP effects could indicate domain-specialised neurocognitive substrates for each process, and, therefore, domain-specialised neurocognitive substrates in language and music domains.

3.2.5. Experiment 2: Expertise effects

To investigate whether English language experience or musical training affected the later-onset classic ERP effects: the N400, the P600 and P300, additional analyses were conducted. The ANOVA designs included two between-subjects measures: English language experience (2 groups: 12 native and eight non-native English speakers) and musical training (2 groups: 10 amateur musicians and ten non-musicians). For descriptive statistics of English language experience for native and non-native speakers, see Table 3.7. For musical training, the cut-off was > 1000 hours of practice time for participants to be categorised as amateur musicians and < 1000 hours of practice time for non-musicians. Descriptive statistics of musical training experience for amateur musicians and non-musicians are in Table 3.8, and the frequency distribution of accumulated practice time is in Appendix F.

Table 3.7 – Descriptive statistics of English language experience for native and non-native speakers, Experiment 2

Measure	Native speakers (*N* = 12)		Non-native speakers (*N* = 8)	
	M (*SD*)	Range	*M* (*SD*)	Range
Onset learning English (years)	0 (0)	0	5.9 (2.6)	3–10
Overall ability rating (1–10)	10 (0)	10	8.4 (1.1)	7–10
% time speaking English	98 (9)	70–100	59 (13)	45–80

Table 3.8 – Descriptive statistics of musical training experience for amateur musicians and non-musicians, Experiment 2

Measure	Amateur musicians (*N* = 10)		Non-musicians (*N* = 10)	
	M (*SD*)	Range	*M* (*SD*)	Range
Number of years practising	10.6 (3.4)	4–14	3.4 (3.8)	1–13
Practice time (hours)	3149 (2364)	1170–8424	178 (230)	40–780
Overall ability rating (1–5)	3.4 (1.0)	2–4	1.5 (0.7)	1–3
Listening per week (hours)	30.0 (22.7)	3–70	15.7 (10.1)	3–35

First, associations between English language experience or musical training experience for behavioural measures were investigated. For the frequency of “Expected” responses in the language task, there was a significant interaction between English language experience and sentence end type $F(3, 48) = 15.61$, $p < .001$, $\eta p^2 = 0.49$. To investigate, pairwise comparisons were made to compare each sentence end type for native and non-native English speakers (e.g., standard words for native speakers with standard words for non-native speakers). After Bonferroni correction, the only difference between native and non-native English speakers was for the related semantic violations ($t(7) = 3.86$, $p = .006$). The native speakers found the related semantic violations less expected ($M = 4\%$, $SD = 3\%$) compared to the non-native speakers ($M = 24\%$, $SD = 15\%$). In the music task, there were no main effects or interactions of either English language experience or musical training experience for the frequency of “Expected” responses. For both language and music tasks, there were no main effects of either English language experience or musical training on reaction time.

Then, mixed ANOVAs were carried out to investigate whether English language experience or musical training affected the N400, P600 or P300 ERP effects. In these analyses, there were no within- or cross-domain effects or interactions associated with either English language experience or musical training (all *p*s > .05).

3.2.6. Experiment 2: Discussion

The behavioural results of Experiment 2 show that the expectancy ratings found in Experiment 1 for the different chord ending types in the music domain are reliable. In both experiments, the tonic (I) chord ending was mostly rated expected, the related harmony was rated “Expected” on about half the trials and “Unexpected” on the other half, and the unrelated harmony was mostly rated as “Unexpected”. For the language task, participants rated the standard sentence to be “Expected” on most of the trials. The other three sentence end types were rated as mostly “Unexpected”. However, the related semantic violations were significantly more “Expected” than the unrelated semantic violations. Therefore, it remained valuable to investigate whether there were differences in ERP effects between related and unrelated semantic violations, as planned.

Experiment 2 revealed clear and distinct ERP effects elicited by different semantic, grammar and harmonic expectancy violations, indicating domain-specialised neurocognitive substrates for these processes in language and music domains. There was an N400 effect for semantic expectancy violations, a P600 effect for grammar expectancy violations and a P300 effect for harmonic expectancy violations. All three effects were significant for both the ANOVA analyses and the cluster-based permutation statistics. The N400 effect was only significant for the more severe semantic expectancy violations (unrelated semantic violations). The P300 effect was significant for both less severe and more severe harmonic expectancy violations (for both related harmony and unrelated harmony chords). However, there was no difference between related and unrelated harmony chords, so a linear effect of harmonic expectancy violation severity was not established.

In the music task, there were two unforeseen clusters identified by the cluster-based permutation statistics. There was an increase in anterior positivity and posterior negativity for the related harmony chord compared to the tonic (I) chord, and unrelated harmony chord types. These effects could reflect task difficulty. Participants appeared to find it more challenging to decide whether the related harmony chords were expected or unexpected compared to the tonic (I) and unrelated harmony chords. This difficulty is reflected in the behavioural responses (for the related harmony chord, participants responded “Expected” about 50% of the time and “Unexpected” about 50% of the time) and reaction time (reaction time was longer for the related harmony chord compared to the tonic (I) and unrelated harmony chords). Previous studies suggest that reaction times can be slower to more surprising stimuli, reflecting additional processing effort (Warrick et al., 1965). The presence of these additional components suggests that effects for expectation and task difficulty may be overlapping in the ERP measurements. This issue is particularly relevant as both the P300 effect and one of the unforeseen clusters are frontally distributed, and the P300 time window (250–400 ms) overlaps with the time course of the cluster, which started at 219 ms and continued until the end of the epoch (800 ms). The spatial and temporal overlaps of these effects could hide or distort the measurement of the expectancy effects of interest.

3.3. Experiment 3: ERP correlates of semantic, grammar and harmonic expectancy violations – an explicit task

3.3.1. Experiment 3: Rationale and aims

Before discussing the results of Experiment 2 in more detail, Experiment 3 is presented. Experiment 3 adopted a task where the expectancy violations were task-irrelevant. For the language task, participants were asked whether the sentence was “Green?” or “Blue?” and answered “Yes” or “No”, and for the music task, participants were asked whether the chord progression was played by the “Piano?” or “Organ?” and answered “Yes” or “No”. Crucially, as task difficulty did not differ between the three chord progression end types, it was predicted that the unforeseen clusters, thought to be associated with differences in task difficulty between the three chord progression end types in the music task in Experiment 2 would not be found in Experiment 3. Based on the results of Experiment 2, it was predicted that an N400 would be found for semantic expectancy violations, a P600 for grammar expectancy violations and a P300 for harmonic expectancy violations.

3.3.2. Experiment 3: Method

Ethics

The University Teaching and Research Ethics Committee at The University of St Andrews approved this study, code PS13256 (see Appendix P).

Participants

Twenty-one university students volunteered to participate in this study. The data of one participant was not used for analysis due to technical issues during EEG recording. Therefore, the final *N* = 20 (6 male, Age: *M* = 27 years, *SD* = 5 years, range = 19–41 years). Descriptive statistics of participants' English language and musical training experience are in Table 3.9. Each testing session took 120 minutes, and participants were reimbursed £10. People could not sign up to participate in the study if they had participated in Experiments 1 or 2.

Table 3.9 – Descriptive statistics of participants' English language and musical training experience, Experiment 3.

Experience	Measure	*M* (*SD*)	Range
English language	Native English speaking	Native: 9, Non-native: 11	
	Onset (age in years)	4.3 (5.3)	0 –18
	English ability (rating 1–10)	9.2 (1.2)	6–10
	% English spoken in the past year	76.4 (21.0)	40–100
Musical training	Number of years practising	8.5 (7.9)	0–25
	Practice time (hours)	2600 (4007)	0–13728
	Overall ability rating (1–5)	2.9 (1.0)	1–4
	Listening per week (hours)	16.4 (15.5)	2–50

Apparatus

See section 3.2.1 Experiment 2: Method (Apparatus).

Materials

Sentences

In the language task, 192 sentences were presented. The sentences in Experiment 3 were the same as those in Experiment 2, apart from one difference. The only difference was that they were now presented in two colours, green and blue. The 96 (24 sentence stems x 4 sentence end types: standard word, related semantic violation, unrelated semantic violation and grammar expectancy violation) sentences were each presented twice, once in green and once in blue. For each of the sentences (in both green and blue), participants were asked either “Green?” or “Blue?”, to which they answered “Yes” or “No”.

Chord progressions

In the music task, 144 chord progressions were presented. The chord progressions in Experiment 3 were the same as in Experiment 2. These 72 chord progressions (24 chord progression stems x 3 chord progression end types: tonic (I), related harmony and unrelated harmony) were each repeated twice, once played by the piano and once played by the organ. For each of the chord progressions (played by piano and organ), participants were asked either “Piano?” or “Organ?” For an illustration of the three chord progression types, refer back to Figure 2.1

Additional materials

In previous studies, encourage participants to read the sentences, they have been told they would answer questions about them at the end of the experiment (Goregliad Fjaellingsdal et al., 2016; Kamp et al., 2015; Kutas & Hillyard, 1983). The same approach was used in the current experiments for the implicit language task.

Five questions were used to check whether participants had paid attention to the sentences:

1. What did the cleaner dust?
2. What animal did George have?
3. Who proposed?
4. What did the father carve?
5. Who was scared of getting bitten?

See section 3.2.1 Experiment 2: Method (Materials) for details of the English Language Experience Questionnaire and section 2.2.2 Experiment 1: Method (Materials) for details of the Musical Training Questionnaire.

Procedure

The procedure of Experiment 3 matched the procedure of Experiment 2, with only one difference: the participants' task. Participants were not asked whether the final chord or word was "Expected?" or "Unexpected?" (as in Experiment 2). Instead, they were asked whether the chord progression was played by "Piano?" or "Organ?" and whether the sentence was presented in "Green?" or "Blue?" (see Figure 3.9). These tasks were chosen because the expectancy violation was task-irrelevant. Thus, there should be no differences in task difficulty between different end types, and it was predicted that only ERP effects that are related to expectancy violations would be elicited.

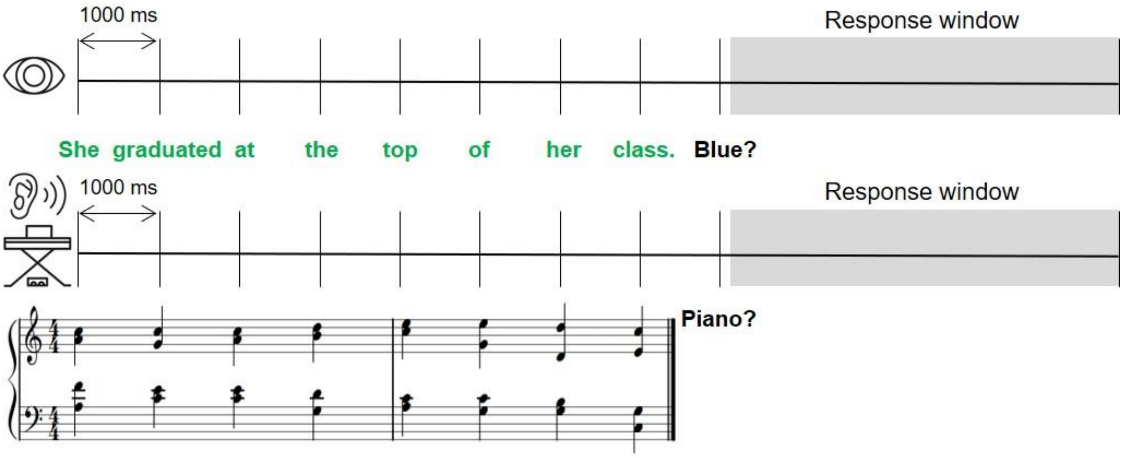


Figure 3.9 – The timing of stimuli presentation for the language task (above) and the music task (below). The standard word and tonic (I) chord end types are illustrated in the examples. For the language example, the correct answer would be: “No”, and for the music example: “Yes”. One word or chord was presented every 1000 ms. The response window (100–5000 ms after question onset) is marked with a shaded area.

As in Experiment 2, participants chose whether they answered “Yes” with the right hand and “No” with the left (*N* = 11), or vice versa (*N* = 9), the order of language task and music task was counterbalanced between participants, and there were two practice blocks involving 12 sentences and 12 chord progressions. Each practice block was completed immediately before the relevant main block. Immediately after the language task, participants were asked to answer the five questions about the sentences to check that they paid attention to the sentences. After both tasks were completed, participants filled in the English Language Experience Questionnaire and the Musical Training Questionnaire.

Analysis

Only trials with correct responses given between 100–5000 ms after the onset of the question prompt were considered for analysis. Incorrect responses were not included in the analysis because the tasks were relatively easy, and so incorrect responses may indicate lapses in attention. Furthermore, previous studies report ERP effects associated with errors (Jentzsch et al., 2014; Picton et al., 2012), which could obscure the current effects of interest. As in Experiment 2, trials were also excluded with artefact rejection using the moving peak-to-peak window method after ICA correction in ERPLAB. Table 3.10 shows how many trials were analysed after the inclusion criteria were applied. For all end types, there were, on average, more than 39 trials in the final dataset. Following Luck's (2014) suggestion that between 10–25% exclusion of 30–40 trials per condition with 20 participants is an acceptable signal-to-noise ratio for late-onset ERP components, the current dataset had an acceptable signal-to-noise ratio for analysis.

Table 3.10 – Mean (*SD*) number of trials included in the analysis after the inclusion criteria were applied in Experiment 3. The maximum possible number of trials for each end type was 48.

		Mean number of trials (*SD*)	
Task	End type	Behaviour	Artefact rejection
Music	Tonic (I)	46.1 (3.4)	41.4 (8.2)
	Related harmony	45.4 (4.5)	39.3 (10.0)
	Unrelated harmony	45.3 (3.8)	41.0 (8.1)
Language	Standard word	47.7 (0.6)	42.3 (6.9)
	Related semantic	47.5 (0.9)	41.6 (6.9)
	Unrelated semantic	47.4 (0.9)	42.5 (7.8)
	Grammar violation	47.1 (0.9)	42.6 (6.6)

The same a-priori time windows and electrodes identified for ANOVA analysis by the cluster-based permutation analysis results in Experiment 2 were analysed in Experiment 3, to facilitate direct comparison between experiments, see Table 3.11 and Figure 3.10. Full details of the ERP analysis are in section 3.2.1 Experiment 2: Method (Analysis).

Table 3.11 – A-priori chosen electrodes and time windows for analysis of N400, P600 and P300 components

ERP component	Electrode	Time (ms)	Violation type
N400	Pz	300–450	Semantic
P600	Cz	500–800	Grammar
P300	Fz	250–400	Harmonic

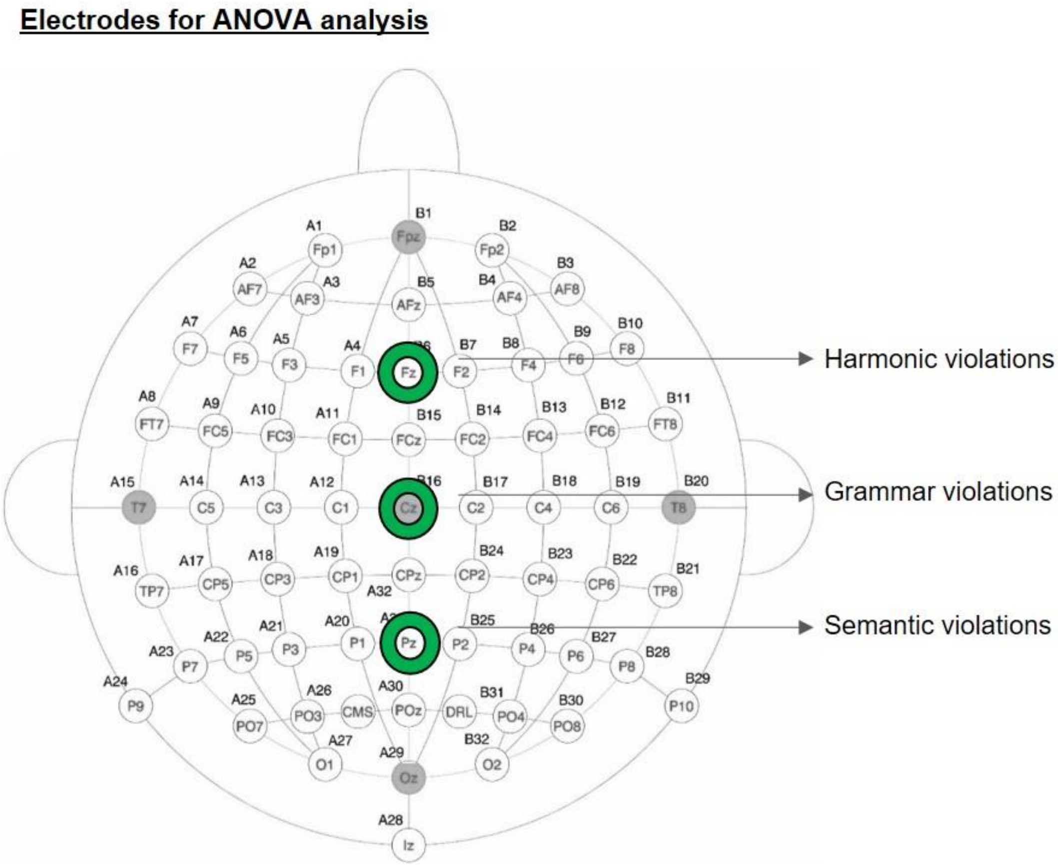


Figure 3.10 – The positions of the three scalp electrodes selected for ANOVA analysis based on the clusters identified in Experiment 2: Fz for harmonic violations, Cz for grammar violations and Pz for semantic violations.

3.3.3. Experiment 3: Behaviour results

On average, participants answered 75% (*SD* = 21%) of the questions about the sentences correctly, providing some evidence that they had read the sentences, and were not just looking at the colour of the sentences. Four within-subjects ANOVAs were carried out to investigate: whether end types affected the percentage of errors participants made on A) the language task and B) the music task, and whether end types affected reaction time in C) the language task or D) the music task. For the language task, there were 4 levels of sentence end type (standard, related semantic violation, unrelated semantic violation and grammar expectancy violation) and for the music task, there were 3 levels of chord progression end type (tonic (I), related harmony and unrelated harmony).

Percentage of errors

For both the language and music task, there was no effect of end type on the percentage of errors (*p* > .05), see Figure 3.11.

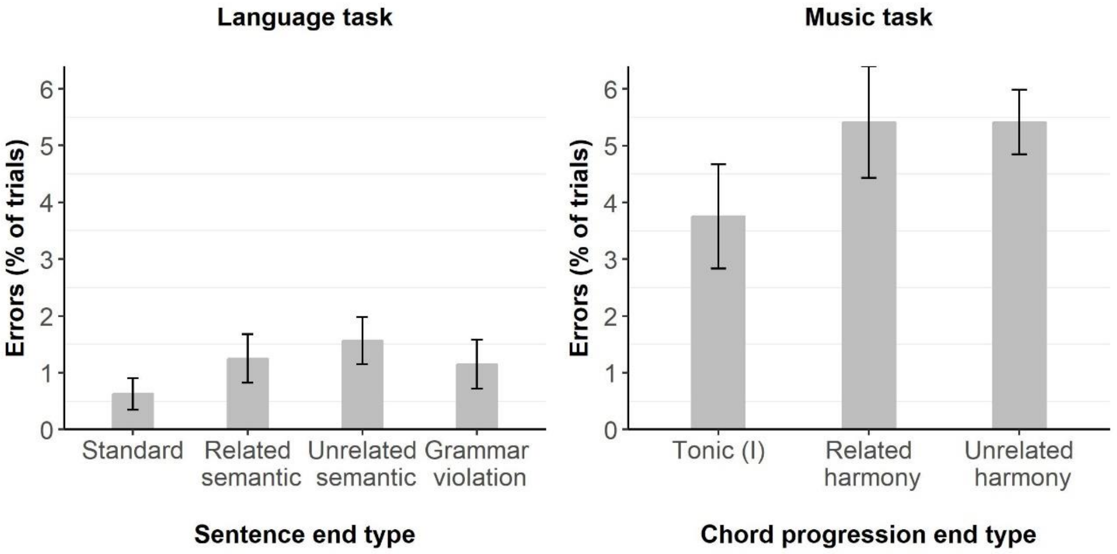


Figure 3.11 – Mean % of errors for each end type in the language task (left) and the music task (right). Within-subject *SE* bars are included

Reaction time

There was no effect of sentence end type on reaction times for the language task (*p* > .05), see Figure 3.12, left panel. However, for the music task, there was an effect of chord progression end type on reaction times *F*(2, 38) = 8.98, *p* = .001, ηp^2 = 0.32, see Figure 3.12, right panel. Pairwise comparisons showed that the tonic (I) chords were responded to faster *(M* = 866 ms, *SD* = 280 ms) than the related harmony chords *(M* = 923 ms, *SD* = 285 ms, *p* = .003) and unrelated harmony chords *(M* = 900 ms, *SD* = 273 ms, *p* = .029). This result could indicate the effect of interest – an expectedness effect. Although the expectancy violations were not made explicit by the task, reaction time may have been slowed by an additional processing effort with the harmonic violation end types, compared to the anticipated (tonic, I) chords.

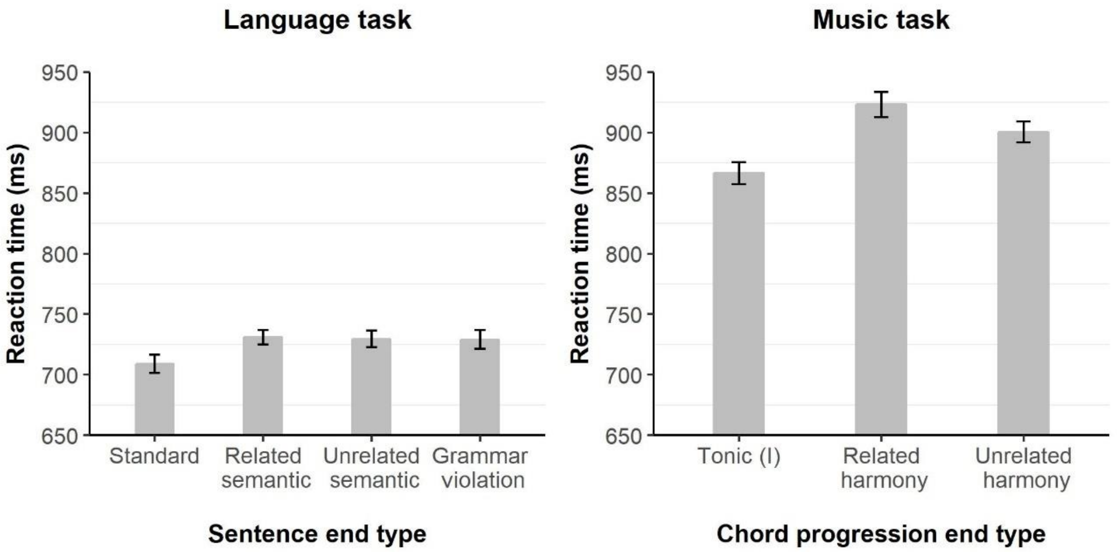


Figure 3.12 – Mean reaction time (ms) for each end type in the language task (left) and the music task (right). Within-subject *SE* bars are included.

Comparison between language task and music task

Two *t*-tests were conducted to compare participants' performance between the language and music tasks. First, the percentage of errors was collapsed over end types for each task. Participants made fewer errors in the language task *(M* = 1.1%, *SD* = 0.9%) than the music task *(M* = 4.9%, *SD* = 7.4%), *t*(19) = 2.23, *p* = .038, *d* = 0.71. A second *t*-test compared reaction times between the two tasks. Across end types, participants responded significantly faster in the language task *(M* = 725 ms, *SD* = 164 ms) than the music task *(M* = 897 ms, *SD* = 164 ms), *t*(19) = 3.33, *p* = .004, *d* = 0.77. These results suggest that participants found it easier to decide if sentences were green or blue than if the chord progressions were played by the piano or organ.

3.3.4. Experiment 3: ERP results

Language domain ERP results: Semantic expectancy violations

Within-subjects cluster-based permutation statistics with three levels of sentence end type (standard words, related semantic violations and unrelated semantic violations) identified a significant cluster (mass = 1024, *p* =.006), in centroparietal areas (P1, P3, PO3, Pz, C4, CP4, CP2, P2, P4, PO8, PO4, O2) from 352–531 ms, see Figure 3.13. The temporal peak was 391 ms, and the spatial peak was at PO4. No additional clusters were identified. A within-subjects ANOVA showed an effect in mean amplitude at Pz in the N400 time window (300–450 ms), *F*(2, 38) = 4.91, *p* = .013, ηp^2 = 0.21. There was an increase in negativity for the unrelated semantic violations (*M* = –0.61 µV, *SD* = 0.45 µV) compared to the standard words (*M* = 0.24 µV, *SD* = 0.32 µV, *p* = .043). There were no significant pairwise comparisons for the related semantic violations (*p*s < .05).

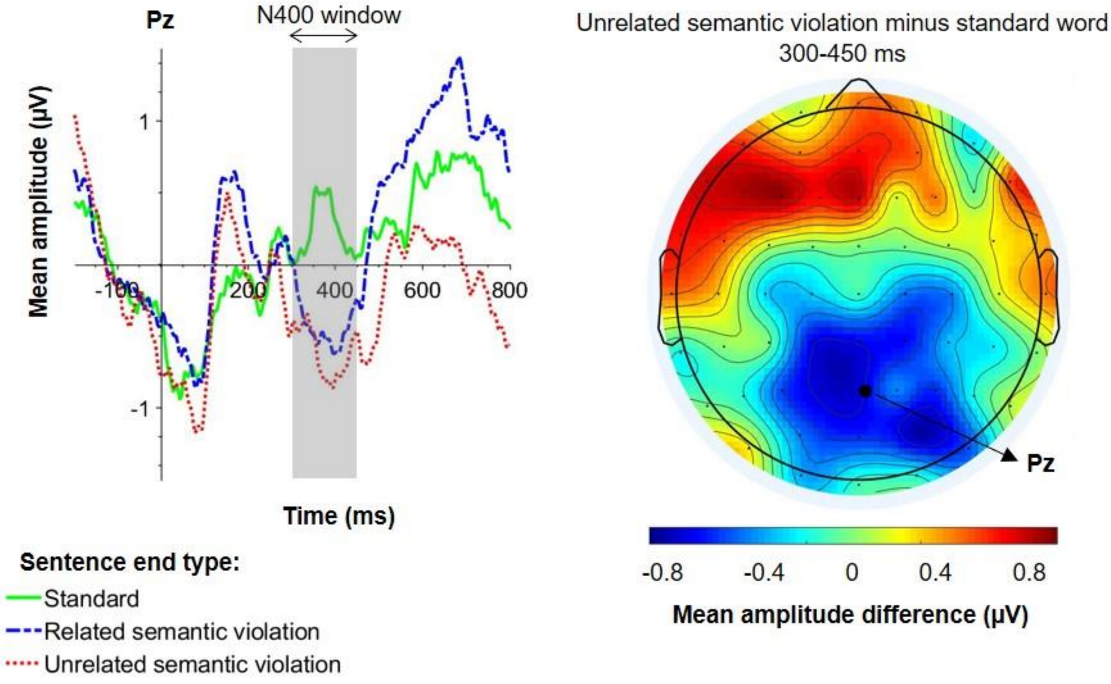


Figure 3.13 – N400 semantic expectancy violation graphs. Left: Mean amplitude (µV) at Pz between –200 to +800 ms for the three sentence end types. Right: Scalp topography map of the mean amplitude difference (µV) between 300–450 ms after the final (target) word's onset.

Language ERP results: Grammar expectancy violations

Within-subjects cluster-based permutation statistics with two levels of sentence end type (standard words and grammar expectancy violations) show a corresponding cluster (mass = 3392, *p* <.001) in central areas (CP3, CP1, P1, P3, Pz, C4, CP4, CP2, P2, P4) from 413 ms until the end of the tested epoch (800 ms), see Figure 3.14. The temporal peak was 563 ms, and the spatial peak was at CP4. No additional clusters were identified. Within-subjects ANOVAs showed that sentence end type affected the mean amplitude at Cz in the P600 time window (500–800 ms) *F*(1, 19) = 4.82, *p* = .041, ηp^2 = 0.20. The mean amplitude was more positive for grammar expectancy violations *(M* = –1.65 µV, *SD* = –0.50 µV) compared to the standard words *(M* = –2.60 µV, *SD* = 0.60 µV).

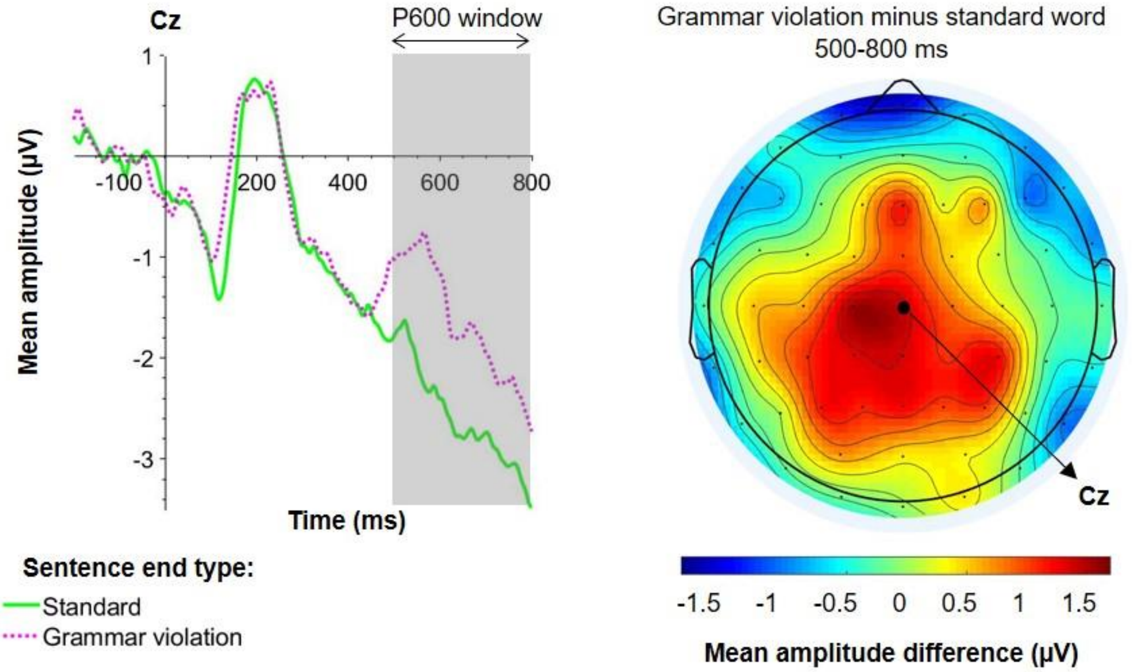


Figure 3.14 – P600 grammar expectancy violation graphs. Left: Mean amplitude (µV) at Cz between –200 to +800 ms for the two sentence end types: standard word and grammar expectancy violation. Right: Scalp topography map of the mean amplitude difference (µV) between 500–800 ms after the final (target) word's onset.

Music ERP results: Harmonic expectancy violations

Within-subjects cluster-based permutation statistics with three levels of chord progression end type (tonic (I), related harmony and unrelated harmony chords) identified a cluster (mass = 1284, *p* = .022) with frontal distribution (F1, F3, FC3, FC1, C3, Fz, F2, F4) between 192–438 ms, with a temporal peak of 305 ms and a spatial peak of FC1, see Figure 3.15. No additional clusters were identified. The ANOVA showed that chord progression end type affected the mean amplitude at Fz during the P300 time window, between 250–400 ms, *F*(2, 38) = 5.41, *p* = .009, $\eta p^2 = 0.22$. It was more positive for the unrelated harmony chords *(M* = 0.66 µV, *SD* = 0.38 µV, *p* = .011) compared to the tonic (I) chords *(M* = –0.50 µV, *SD* = 0.29 µV). Pairwise comparisons showed no difference between related harmony *(M* = –0.23, *SD* = 0.34) and tonic (I) chords (*p* = 1.00), but showed a difference between related and unrelated harmony chords (*p* = .045).

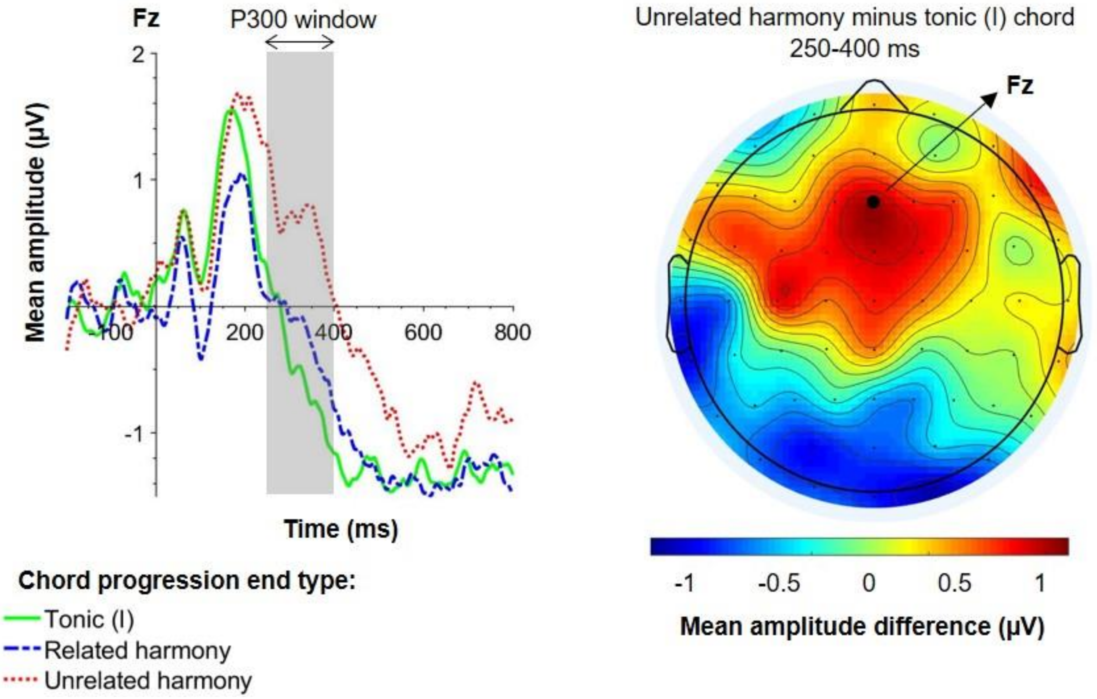


Figure 3.15 – P300 harmonic expectancy violation graphs. Left: Mean amplitude (µV) at Fz between –200 to +800 ms for the three chord progression end types: tonic (I), related harmony and unrelated harmony. Right: Scalp topography map of the mean amplitude difference (µV) between 250–400 ms after the final (target) chord's onset.

3.3.5. Experiment 3: Test of distinct ERP effects

The results discussed in the previous section show that there were clusters associated with the N400 effect for semantic expectancy violations, the P600 effect for grammar expectancy violations and the P300 effect for harmonic expectancy violations. Nieuwenhuis et al. (2011) suggest that to argue that two ERP components, elicited by different stimuli, are distinct, it is necessary to establish that for each effect, there is no comparable effect for the other stimulus type. Therefore, ANOVA analyses were run that tested each time window and electrode of interest with the other types of expectancy violations. The results of all these analyses were not significant (all *ps* > .05), see Appendix H for full details. Altogether, the evidence suggests that there are distinct ERP effects for semantic, grammar and harmonic expectancy violations. These distinct ERP effects could indicate domain-specialised neurocognitive substrates for each process and, therefore, domain-specialised neurocognitive substrates in language and music domains.

3.3.6. Experiment 3: Expertise effects

To investigate whether native English speaking or musical training moderated the N400, P600 and P300 effects, additional mixed ANOVA analyses were conducted. The ANOVA designs included two between-subject measures: English language experience (2 groups: 11 native and nine non-native English speakers, see Table 3.12 for descriptive statistics) and musical training (2 groups: 9 amateur musicians and 11 non-musicians). For descriptive statistics of musical training experience for amateur musicians and non-musicians, see Table 3.13. The cut-off was > 1000 hours of practice time for participants to be categorised as amateur musicians, and < 1000 hours of practice time for non-musicians. For the frequency distribution of accumulated practice times, see Appendix F.

Table 3.12 – Descriptive statistics of English language experience for native and non-native speakers, Experiment 3

Measure	Native speakers (*N* = 11)		Non-native speakers (*N* = 9)	
	M (*SD*)	Range	*M* (*SD*)	Range
Onset learning English (years)	0 (0)	0	7.8 (4.9)	2–18
Overall ability rating (1–10)	10 (0)	10	8.5 (1.2)	6–10
% time speaking English	88 (16)	60–100	70 (19)	45–95

Table 3.13 – Descriptive statistics of musical training experience for amateur musicians and non-musicians, Experiment 3

Measure	Amateur musicians (*N* = 9)		Non-musicians (*N* = 11)	
	M (*SD*)	Range	*M* (*SD*)	Range
Number of years practising	15.6 (3.8)	10–20	2.64 (2.0)	1–7
Practice time (hours)	5564 (4481)	1612–13728	193 (233)	452–832
Overall ability rating (1–5)	3.7 (0.7)	2–4	2.2 (0.8)	1–3
Listening per week (hours)	16.4 (10.4)	4–35	9.0 (6.5)	2–22

First, to investigate whether English language experience or musical training affected behavioural results, mixed ANOVAs were carried out. There were no main effects of English language experience or musical training on either the percentage of participants' errors or reaction time (all *ps* > .05). Then, English language experience and musical training were added to the ANOVAs for the N400, P600 and P300 ERP effects. In these analyses, there were no within- or cross-domain effects or interactions associated with either English language experience or musical training (all *p*s > .05).

3.3.7. Experiments 2 and 3: Results comparison

To investigate whether there were statistically significant differences for the N400, P600 and P300 ERP effects between Experiments 2 and 3 (and therefore, between when the expectancy violations were task-relevant and task-irrelevant), additional ANOVAs were carried out, adding a 2 level between-subject factor of the experiment (Experiment 2 and Experiment 3) to the previous designs. For the N400 and P600 effects, there were no interactions between sentence end type and experiment ($F(2, 76) = 1.27$, $p = .286$, $\eta p^2 = 0.03$ and $F(1, 38) = 0.83$, $p = .370$, $\eta p^2 = 0.02$, respectively). For the P300 effect, there was a non-significant trend for an interaction between chord progression end type and experiment $F(2, 76) = 2.77$, $p = .069$, $\eta p^2 = 0.07$. This trend is driven by the results for the related harmony chords. While there was an increase in positive amplitude for the related chord compared to the tonic (I) in Experiment 2 ($p = .004$), there was no such effect in Experiment 3 ($p = 1.00$, see above). In fact, the P300 effect was lacking for the related harmony chords to the extent that there was a significant P300 for unrelated harmony chords compared to related harmony chords ($p = .045$, see above). Therefore, while the language domain violations were not affected by whether the expectancy violation was task-relevant or task-irrelevant, the task-driven saliency of the harmonic expectancy violations appears to have meaningfully affected P300 effect for the less severe (related harmony) harmonic expectancy violations.

3.3.8. Experiments 2 and 3: Discussion

Aim 1: Is there evidence of distinct ERP effects?

The main aim of Experiments 2 and 3 was to investigate whether there is reliable evidence of distinct ERP effects associated with semantic and grammar expectancy violations in the language domain and harmonic expectancy violation in the music domain. These experiments identified robust ERP components elicited by each violation-type. In both experiments, the N400 effect was found for semantic violations, the P600 effect was found for grammar violations, and the P300 effect was found for harmonic violations. Crucially, there was no evidence of these ERP effects occurring for any other violation types and the scalp topographies for each effect were distinct (the N400 was centroparietal, the P600 was central and the P300 was frontal). These distinct ERP effects could indicate domain-specialised neurocognitive substrates for these processes, and therefore, could indicate separable neurocognitive substrates between language and music domains. In this section, these ERP effects are discussed in turn for semantic, grammar and harmonic expectancy violations.

Semantic expectancy violations

The N400 effect was elicited by semantic expectancy violations in both Experiments 2 and 3, confirming that the semantic expectancy violation stimuli created for this thesis (in Chapter 2) elicit the classic N400 effect (Gutierrez et al., 2012; Kamp et al., 2015; Kutas & Federmeier, 2010; Kutas & Hillyard, 1983; Moreno & Vázquez, 2011; Nigam et al., 1992; Rommers et al., 2013; Tiedt et al., 2020). This N400 effect is spatially and temporally equivalent in Experiments 2 and 3 (when expectancy violations were task-explicit or task-implicit), It had a similar, centroparietal, scalp topography in both experiments. These characteristics fit with N400 effects reported in both visual and auditory modalities in the literature, which are highly similar (Kutas & Federmeier, 2000). Some previous studies have suggested that the N400 is more pronounced in the left hemisphere (Palolahti et al., 2005). However, most report a non-lateralised scalp topography, and the N400 in the current experiments fits with most of the previous findings. The cluster-based statistics reinforce that the chosen a-priori time window to investigate the N400 of 300–450 ms after the final (target) word's onset was a suitable choice, as the corresponding significant clusters mainly fall within this time window in both experiments (297–422 ms in Experiment 2 and 352–531 ms in Experiment 3). Overall, the highly comparable N400 effects between Experiments 2 and 3 suggest two things. First, it is a robust and replicable effect. Second, the task-driven saliency of semantic expectancy violations does not substantially impact the neurocognitive substrates of semantic violation processing, measured by the N400 effect.

In both Experiments 2 and Experiment 3, the N400 effect was only significant for unrelated semantic violations, not for the less severe related semantic violations. However, the descriptive statistics, see Figure 3.5 and Figure 3.13, indicate that there may have been a small N400 effect for the related semantic violations in both experiments. If so, it was not large or sustained enough to detect with the chose a-priori time window and mean amplitude method. Future studies aiming to investigate ERP effects for less severe semantic expectancy violations could consider more focal a-priori time windows.

Critically, the N400 was not elicited by grammar or harmonic expectancy violations. Therefore, the evidence is consistent with two hypotheses. First, that semantic processing relies on separable neurocognitive substrates from grammar processing, as in the Faculty of Language model proposed by Hauser et al. (2002). This is consistent with previous evidence that reports that N400s do not occur when grammar expectancy violations are most prominent, and therefore they do not seem to reflect grammar processing (Kim & Osterhout, 2005). Second, that semantic processing relies on separable neurocognitive substrates from harmony processing. This current finding supports previous findings that musical stimuli do not elicit N400 effects (Besson & Faïta, 1995; Besson & Macar, 1987; Miranda & Ullman, 2007; Paller et al., 1992), and support a model that suggests that neurocognitive substrates involved in semantic processing are an example of separable mechanisms between language and music domains (Brown et al., 2006).

Grammar expectancy violations

The P600 effect was elicited by the grammar expectancy violations in both Experiments 2 and 3, confirming that the grammar expectancy violations created for this thesis elicit a classic P600 effect (Hahne et al., 2012; Mehravari et al., 2015; Osterhout & Holcomb, 1992). The P600 effect had similar characteristics in Experiments 2 and 3, such as scalp topography and latency. Therefore, it does not appear to have been affected by the task-driven saliency of grammar expectancy violations. This idea is supported by the results of Balconi and Pozzoli (2005). They also found that the amplitude and latency of the P600 are not affected by whether grammar expectancy violations are task-relevant or task-irrelevant, or by whether they are presented in the visual or the auditory modality. In both experiments, cluster-based permutation statistics showed that grammar expectancy violation effects begin at around 400 ms after the target stimulus onset and peak around 600 ms. In both experiments, the P600 cluster effect was significant until 800 ms, which was the end of the tested epoch. Consequently, it might be more suitable for future studies to test a wider epoch, e.g., starting at 400 ms and lasting until 1000 ms after stimulus onset. Such an extended time window could allow its full temporal extent to be examined.

Crucially, the P600 was not elicited by semantic or harmonic expectancy violations. Therefore, the current evidence is consistent with two hypotheses. First, that grammar processing relies on separable mechanisms from semantic processing, as suggested by previous evidence (Kim & Osterhout, 2005) and the Faculty of Language model, proposed by (Hauser et al., 2002). Second, grammar processing relies on separable mechanisms from harmony processing (Frisch et al., 2003; Osterhout, 1999). It seems that the “P600 effect” reported for harmonic expectancy violations by Patel et al. (1998) may have represented later-onset components, as they occurred 1100 ms after the harmonic expectancy violation onset. These later-onset components and their comparison to grammar violation ERP effects require further, separate investigation.

Harmonic expectancy violations

A P300 effect was elicited by the harmonic expectancy violations in both Experiments 2 and 3, confirming that the harmonic expectancy violations systematically created in Experiment 1 elicit a P300 effect. These P300 effects had similar, frontal, scalp topography Experiments 2 and 3, and appear to indicate the same neurocognitive substrates. The cluster-based permutations statistics suggest that to investigate the P300 effect, a wider time window of 200–400 ms might be suitable than the current 250–400 ms because the clusters corresponding to the P300 fell within these times (211–461 ms in Experiment 2 and 192–438 ms in Experiment 3). Previous studies have reported similar P300 effects for harmonic violations (Beisteiner et al., 1999; Carrión & Bly, 2008; Janata, 1995; Loehr et al., 2013), and the current findings support that the P300 is a robust effect for harmonic violations.

In Experiment 2, task demands appeared to be reflected in unforeseen effects identified by the cluster-based permutation statistics. These task difficulty effects were eliminated in Experiment 3 the task was irrelevant to the expectancy violations and the task demands were the same between the three chord progression types. Thus, the evidence supports the interpretation that the additional effects in Experiment 2 are due to the task demands, as proposed in the section 3.2.6, rather than the expectancy of the stimuli. Previous studies also report that task demands can elicit additional ERP effects (Ellison et al., 2015). Together, these findings illustrate the importance of testing task designs in ERP studies.

Comparisons between Experiments 2 and 3 suggest that the P300 can be affected by the saliency of harmonic expectancy violations in the case of less severe harmonic expectancy violations. In Experiment 2, when the task made the harmonic expectancy violations more salient, there was a significant increase in P300 amplitude for both the related and unrelated harmony chords, compared to the tonic (I) chord. However, in Experiment 3, when the expectancy violation was task-irrelevant, the P300 effect was only found for the unrelated harmony chord, and not for the related harmony chord. Moreover, a P300 effect was so lacking for the related harmony chords that there was both no difference between the tonic (I) and related harmony chords and a significant P300 effect for the unrelated harmony chords compared to the related harmony chords. Previous research suggests that the P300 effect is larger when attention is drawn to the harmonic expectancy violation by the task (Polich, 2007) and that P300s can be most likely to occur when the task makes the harmonic expectancy violations explicit (Friedman et al., 2001). The current experiments support that this is the case for less-severe harmonic expectancy violations, but perhaps when harmonic expectancy violations are severe enough, they can elicit P300 effects of similar magnitude whether the expectancy violation is task-relevant or not.

There was no evidence of a P300 effect being elicited by semantic or grammar expectancy violations and, therefore, the evidence is consistent with the hypotheses that harmony processing relies on separable neurocognitive substrates from semantic processes (Besson & Faïta, 1995; Besson & Macar, 1987; Miranda & Ullman, 2007; Paller et al., 1992) and grammar processes (Frisch et al., 2003; Osterhout, 1999).

Aim 2: Comparing task difficulty between language and music tasks

Experiments 2 and 3 aimed to balance task difficulty between language and music tasks, as Magne et al. (2005) suggest that this aids neural comparisons. The aim to balance task difficulty between domains appears to have been met in Experiment 2. There were no task-based differences in reaction time, and reaction time is an indicator of task difficulty (Palmer et al., 1994; Warrick et al., 1965). On the other hand, task difficulty does not seem to be balanced between language and music tasks in Experiment 3. There were fewer errors and quicker reaction times in the language task than the music task, suggesting that participants may have found the language task easier than the music task. Despite this, the results still provide strong evidence for domain-specialised neurocognitive substrates in language and music domains, as the ERP effects had similar characteristics for semantic, grammar and harmonic expectancy violations in Experiment 2 (when task difficulty appears to be balanced between language and music tasks) and Experiment 3 (when task difficulty appears to be unbalanced).

Aim 3: Assessing task paradigms

Although the characteristics of the N400 (for unrelated semantic violations), P600 and P300 (for unrelated harmony chords) effects were not statistically different between the explicit task and the implicit task, one could argue that they were less clear both in terms of scalp topography and temporal characteristics and that there was more noise in the EEG signal for the implicit task compared to the explicit task (see Figure 3.16). This finding is in line with those of Erlbeck et al. (2014) and Hahne & Friederici, 2002, who found (respectively) that the N400 and P600 were smaller for implicit compared to explicit tasks. In the current study, it is unfortunately not possible to tell whether the reduced clarity of the effects is due to task-induced saliency of the violations or participants’ focus on reading the sentences and listening to the chord progressions.

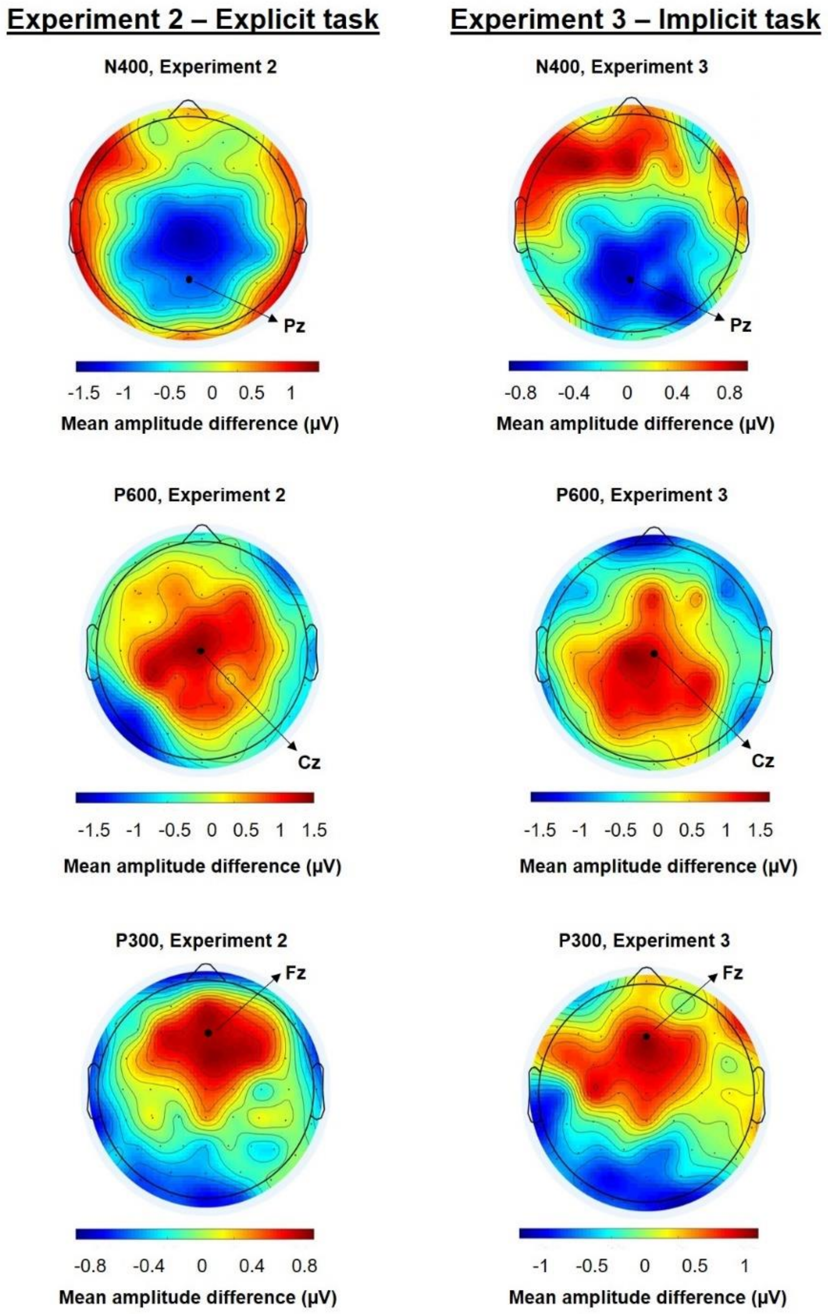


Figure 3.16 – A comparison of effects between Experiment 2 (left) and Experiment 3 (right). These maps are from the figures presented in the results sections of Experiments 2 and 3 and are organised here to facilitate comparison of ERP effects between experiments. Please note: these scalp maps are on different scales.

Despite being told that there would be questions about the sentences to test that they had read them, the participants may have been less motivated to read the sentences all the way through when the task was to answer whether the sentence was written in blue or green compared to when the task was to answer whether the final word was expected or unexpected in the context of the rest of the sentence. Therefore, there appears to be a trade-off of the task-type. When the task was explicit in relation to the expectancy violations, ERP effects were clearer. However, were additional ERP effects in the music task that could interfere with measurements of effects of interest. When the task was implicit in relation to the expectancy violations, ERP effects were less clear, but there were no additional effects, that could potentially obscure the effects of interest.

Moreno and Vázquez (2011) showed that the amplitude of the N400 effect for semantic violations was not affected when participants were warned that a semantic expectancy violation was about to occur. In the current experiments, after the first few trials, participants were likely to be anticipating that there might be expectancy violations. However, the N400, P600 and P300 effects still occurred in both Experiments 2 and 3. Therefore, similar to what Moreno and Vázquez (2011) proposed, these effects could indicate neurocognitive substrates related to mechanisms for processing expectancy violations, regardless of whether their occurance is anticipated or not.

Aim 4: Testing the expectancy of the language violation stimuli

The stimuli for semantic and grammar expectancy violations were created based on a rich literature investigating these types of expectancy violations. Because of this rich literature with tried and tested methods, it was not deemed necessary to run an additional behavioural experiment to assess the expectancies of different semantic and grammar expectancy violations before conducting the ERP experiments. The results of Experiment 2 show clear expectancy ratings for the standard sentence (rated mostly "Expected") and the unrelated semantic and grammar expectancy violations (rated mostly "Unexpected"). Although the less severe, related semantic violations were rated as more "Expected" than the unrelated semantic violations, they were still rated as mostly unexpected on the whole. The frequency of "Expected" responses was *M* = 13% (*SD* = 9%). The related semantic endings may have been rated mostly unexpected because the sentences created strong expectancies for the standard endings, which were used for the standard words, > 90% (Block & Baldwin, 2010; Bloom & Fischler, 1980). Even though these related violations were mostly rated “Unexpected” in the current experiments, there was no N400 semantic expectancy violation effect for them. Nevertheless, on reflection, it is recommended that future studies aim to separately assess participants’ ratings of expectancy before conducting neural studies, particularly if the expectancy ratings are not collected during the neural study (e.g., if only implicit tasks are used).

Aim 5: Investigating expertise effects

There were no moderating within- or cross-domain expertise effects for either English language experience or musical training, for any of the ERP effects in Experiments 2 or 3. There was one behavioural effect of English language experience on expectancy ratings in the language task. Non-native English speakers rated the related semantic violations as more “Expected” than native English speakers in Experiment 2. This effect highlights the importance of future studies collecting behavioural evidence about how participants in different groups rate the expectancies of the expectancy violations, to aid the interpretation of any neural differences between groups.

However, there were no within- or cross-domain effects of English language experience or musical training on the N400, P600 or P300 effects in the current experiments. This differs to some previous research, which reports within-domain expertise effects of language and music expertise on expectancy violation ERP effects (Anurova & Immonen, 2017; Batterink & Neville, 2013; Steinbeis et al., 2006) and a cross-domain expertise effect of an increased N400 semantic expectancy violation effect in children with formal musical training compared to children without musical training (Dittinger et al., 2017). It is possible that, in the current experiments, the differences in proficiency between native and non-native English speakers and amateur musicians and non-musicians were not distinct enough to detect expertise effects. This idea is discussed in more detail in section 6.1.6.

3.4. Conclusion

The experiments presented in this chapter (Experiments 2 and 3) were designed to test whether ERP effects for semantic, grammar and harmonic expectancy violations are distinct and, therefore, whether they might provide evidence to support the hypothesis that there are domain-specialised neurocognitive substrates for these processes. These experiments are the first to investigate ERP effects associated with semantic, grammar and harmonic expectancy violations for both language and music in a within-subjects design, and with attempts at analogous task designs.

The results suggest that there are distinct ERP effects for semantic (N400), grammar (P600) and harmony (P300) violations. There were no additional expectancy violation effects captured by the cluster-based permutation statistics for any of these processes. Crucially, all ERP effects identified in Experiments 2 and 3 occurred for only one type of violation, and their spatial and temporal characteristics were distinct. As semantics and grammar are major components of the language domain, and harmony is a major element of the music domain, these findings suggest that there could be domain-specialised neurocognitive substrates in language and music domains.

The results of Experiments 2 and 3 also demonstrate the importance of considered task design. When designing future studies, it is necessary to consider the trade-off between tasks where the expectancy violation is task-relevant (as in Experiment 2) compared to where they are task-irrelevant (as in Experiment 3). In the current studies, when the expectancy violations were made explicit by the task, they elicited visually clearer ERP effects. However, for the music task, there were additional ERP effects that appeared to be related to task demands, which may obscure the effects of interest. Both explicit and implicit tasks could be valid choices for future studies, depending on their aims. The rest of the EEG experiments in this thesis adopt the implicit tasks to eliminate the additional task effects that are not the expectancy violation effects of interest. Additionally, the less severe ("related") violations for both the sentences and chord progressions are not used in later experiments, as they did not elicit clear and robust ERP effects.

The ERP effects elicited by the expectancy violations in Experiments 2 and 3 are revisited in Experiments 6 and 7 (in Chapter 5) when it is tested if there are interactions in these effects when each violation type is presented simultaneously with meter violations. These later experiments further test whether the ERP effects reported in the current chapter are specialised for semantic, grammar and harmony processing.

4. Chapter 4 – ERP correlates of meter expectancy violations

4.1. Introduction

In the previous chapter, the hypothesis that there are distinct ERP effects for semantic, grammar and harmonic expectancy violations was investigated. In the current chapter, the case for domain-general ERP effects is investigated, with a focus on meter processing. If there are similar ERP effects for meter expectancy violations in language and music domains, this could strengthen the evidence for domain-general neurocognitive substrates involved in meter processing.

As discussed in Chapter 1 (see section 1.3.3), meter is an important aspect of language and music processing. Previous research suggests that there might be shared neurocognitive mechanisms for meter processing between language and music domains (Jackendoff & Lerdahl, 2006; Jusczyk et al., 1999; Lense & Dykens, 2016). The Dynamic Attending Theory suggests that regular beats enhance temporal expectancies for upcoming events (Jones & Boltz, 1989). These enhanced temporal expectancies seem to increase the efficiency of both language processing (Cason et al., 2015; Jusczyk et al., 1999; Mattys & Samuel, 1997; Schmidt-Kassow & Kotz, 2008) and music processing (Palmer & Krumhansl, 1990; Vuust & Witek, 2014). Additionally, the evidence suggests that meter processing can be improved both by formal language training (Elmer et al., 2010; Kalender et al., 2013) and by formal musical training (Slater et al., 2013; Yates et al., 2017; Zhao et al., 2017).

Despite compelling theories and evidence suggesting domain-general neurocognitive substrates for meter processing in language and music domains, there appear to be no ERP studies aiming to actively compare ERP effects for meter violations between the two domains. Meter violations have been created in different ways for language and music experiments presented in the previous literature, limiting the comparison of their associated ERP effects. Meter violations in the language domain tend to be created by manipulating syllable stress patterns (Cox et al., 2016; Zhang & Zhang, 2019), while interstimulus intervals have been manipulated to create meter violations in the music domain. Despite the regular beat being the driving meter in the

English language (Cutler & Otake, 2002; Pitt & Samuel, 1990; Port, 2003), no previous studies seem to have investigated interstimulus interval violations in the language domain.

Most previous manipulations of meter in the language domain occur in the auditory modality. However, there is compelling evidence to suggest that meter is also important for language received in the visual modality. For example, children who struggle to read also tend to have difficulties with meter perception and the temporal processing of speech (Gordon et al., 2015; Huss et al., 2011; Thomson et al., 2006; Tierney & Kraus, 2013). Moreover, Breen et al. (2019) found an N1 effect for metric incongruities during a reading task. Together, these findings motivate the current experiments to investigate ERP effects associated with meter violations in the language domain in the visual modality. The current experiments, Experiments 4 and 5, aim to investigate whether there are comparable ERP effects for meter violations in language and music domains.

4.2. Experiment 4: Testing the expectancy of meter violations

4.2.1. Experiment 4: Rationale and aims

Previous studies lack a systematic investigation of what meter violations people find reliably unexpected. Zhang et al. (2019) created meter violations by changing the interstimulus interval between the final two chords of chord progressions. The other chords of the chord progression were presented 600 ms apart, and the interstimulus interval between the final two chords was either regular (also 600 ms) or irregular (524 ms). However, the participants were asked to complete a distractor task (to press a key when there was a deviant timbre – when a chord was played with a bassoon sound rather than a piano sound). Therefore, it is unclear whether the participants would reliably categorise the meter violation as unexpected. Establishing whether participants have a conscious experience of the meter violations would aid interpretations of associated ERP effects.

Furthermore, previous research has suggested that delayed chords may not be processed as unexpected events, as delayed endings are common in music (Schmuckler & Boltz, 1994), particularly in jazz and rock (Ashley, 2002) and expressive classical music (Sloboda & Lehmann, 2001). This suggestion has led to some research using only shorter irregular interstimulus intervals between the final two chords as meter violations. For example, Zhang et al. (2019) cite this as their reason for not testing longer interstimulus intervals. However, perhaps it is necessary to investigate this hypothesis actively. Using only shorter interstimulus intervals leaves open the possibility that neural differences between regular and irregular meters are just due to time differences, rather than the effect of interest – meter expectancy violations.

The main aim of Experiment 4 is to identify interstimulus intervals that participants reliably rate as being similarly "Expected" and "Unexpected" in both language and music domains. Participants were presented with sentences and chord progressions that had different interstimulus intervals between the final two words or chords. The interstimulus interval was either regular (the same interstimulus interval as between the rest of the stimuli in the sequence) or irregular. If interstimulus intervals that were rated similarily "Expected" and "Unexpected" in language and music domains were identified, these stimuli could be used in later experiments to compare ERP correlates of meter violations between domains, and the potential confound of differences in the expectancy of the stimuli between domains would be removed.

It was planned that within- and cross-domain effects of musical expertise on meter violation effects would be investigated in Experiment 5. Hence, Experiment 4 also aimed to examine whether participants' expectancy ratings were moderated by musical training[7]. Previous studies report differences in meter ERP effects due to musical training (Habibi et al., 2014; Jongsma et al., 2005; Kober et al., 2015). Therefore, to aid

---

[7] Expertise effects of English language experience were not investigated in Experiments 4 and 5 because there was not an even number of native and non-native English speakers in either Experiment 4 (Native *N* = 5, Non-native *N* = 11) or Experiment 5 (Native *N* = 15, Non-native *N* = 5).

interpretations of any moderating effects of musical training on metric violation ERP effects, it was deemed important to test whether musical training affected participants' expectancy ratings of the metric violations.

4.2.2. Experiment 4: Method

Ethics

The University Teaching and Research Ethics Committee at The University of St Andrews approved this study, code PS13256 (see Appendix P).

Participants

Sixteen university students volunteered to participate in this study (2 male, Age: *M* = 20 years, *SD* = 3 years, range = 17–26 years), see Table 4.1 for descriptive statistics of English language experience. In the recruitment advertisement for this experiment, people were asked to sign up only if they considered themselves either a "musician" or a "non-musician". Participants were categorised into amateur musician and non-musician groups based on their own view of whether they were a "musician" or a "non-musician"[8]. There were eight amateur musicians and eight non-musicians. See Table 4.2 for descriptive statistics of musical training experience for each group, and Appendix F for the frequency of accumulated practice times. Each testing session took 60 minutes, and participants were reimbursed £5.

Table 4.1 – Descriptive statistics of participants' English language experience, Experiment 4.

Experience	Measure	*M* (*SD*)	Range
English language	Native English speaking	Native: 5, Non-native: 11	
	Onset (age in years)	2.4 (5.1)	0–19
	English ability (rating 1–10)	9.8 (0.7)	8–10
	% English spoken in the past year	79 (23)	40–100

---

[8] Group memberships matched with the cut–off chosen after collecting data from all participants, of > 1000 hours of practice time for participants to be categorised as amateur musicians, and < 1000 hours of practice time for non-musicians (see Appendix F).

Table 4.2 – Descriptive statistics of musical training experience for amateur musicians and non-musicians, Experiment 4.

Measure	Amateur musicians		Non-musicians	
	M (*SD*)	Range	*M* (*SD*)	Range
Number of years practising	11.8 (4.9)	4–20	2.75 (2.8)	1–9
Practice time (hours)	3480 (1323)	2080–5524	196 (169)	52–520
Overall ability rating (1–5)	3.4 (0.7)	2–4	2.00 (1.1)	1–4
Listening per week (hours)	20.7 (17.2)	4–55	10.63 (6.4)	1–20

Apparatus

For details of the computer set up and response keys, see section 2.2.2, Experiment 1: Method (Apparatus).

Materials

Sentences

In the language task, 120 sentences were presented. The same 24 sentence stems from Experiments 2 and 3 were used, adapted from Block and Baldwin (2010). However, in Experiment 4, they were only presented with the standard word end type. The 24 sentences were each presented five times. The difference between the five presentations was the interstimulus interval between the penultimate and final word. Compared to the previous interstimulus intervals between words in the sentence, the interstimulus interval between the penultimate and final words was either regular (1000 ms) or irregular: short (500 ms), medium-short (750 ms), medium-long (1250 ms) or long (1500 ms). In addition to the 120 sentences, there were ten practice trials. Two sentences that were not presented in the main experiment were presented with each of the five interstimulus intervals. Figure 4.1 shows an illustration of the five interstimulus intervals, in the order of shortest to longest.

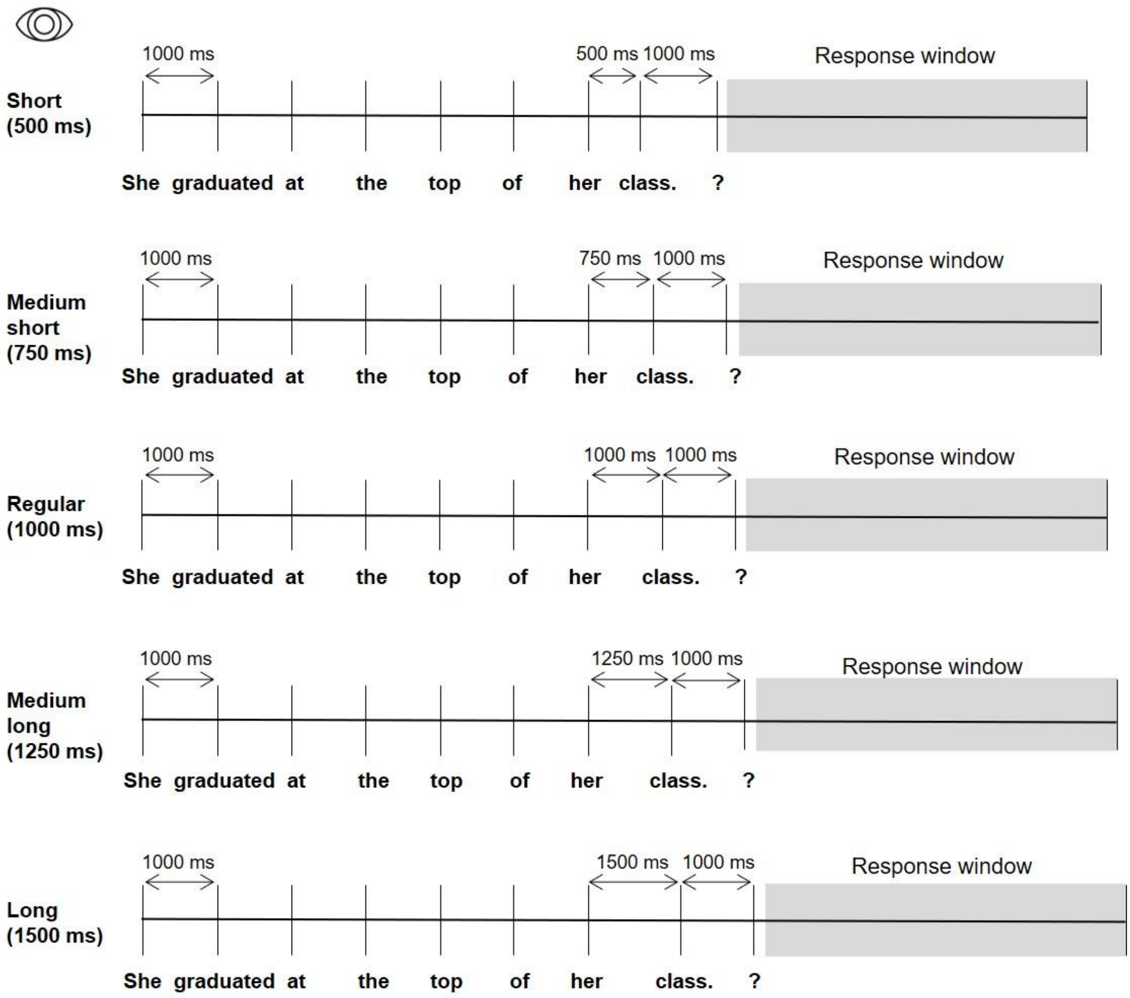


Figure 4.1 – Illustration of the five interstimulus intervals in the language task: short (500 ms), medium-short (750 ms), regular (1000 ms), medium-long (1250 ms) and long (1500 ms). The response window (100–5000 ms after question onset) is marked with a shaded area.

Chord progressions

In the music task, 120 chord progressions were presented. The same 24 chord progression stems from Experiment 2 and Experiment 3 were used, adapted from selected Bach chorales (Bach-Gesellschaft, 1892). In Experiment 4, they were only presented with the tonic (I) end type. The 24 chord progressions were each presented five times. Identical to the language task, the difference between the five presentations in the music task was the interstimulus interval between the penultimate and final chord. The interstimulus interval between the penultimate and final chords was either regular (1000 ms) or irregular: short (500 ms), medium-short (750 ms), medium-long (1250 ms) or long (1500 ms). In addition to the 120 chord progressions presentations, there were ten practice trials, for which two chord progressions that were not presented in the main experiment were presented with each of the five interstimulus intervals. Figure 4.2 shows an illustration of the five interstimulus intervals.

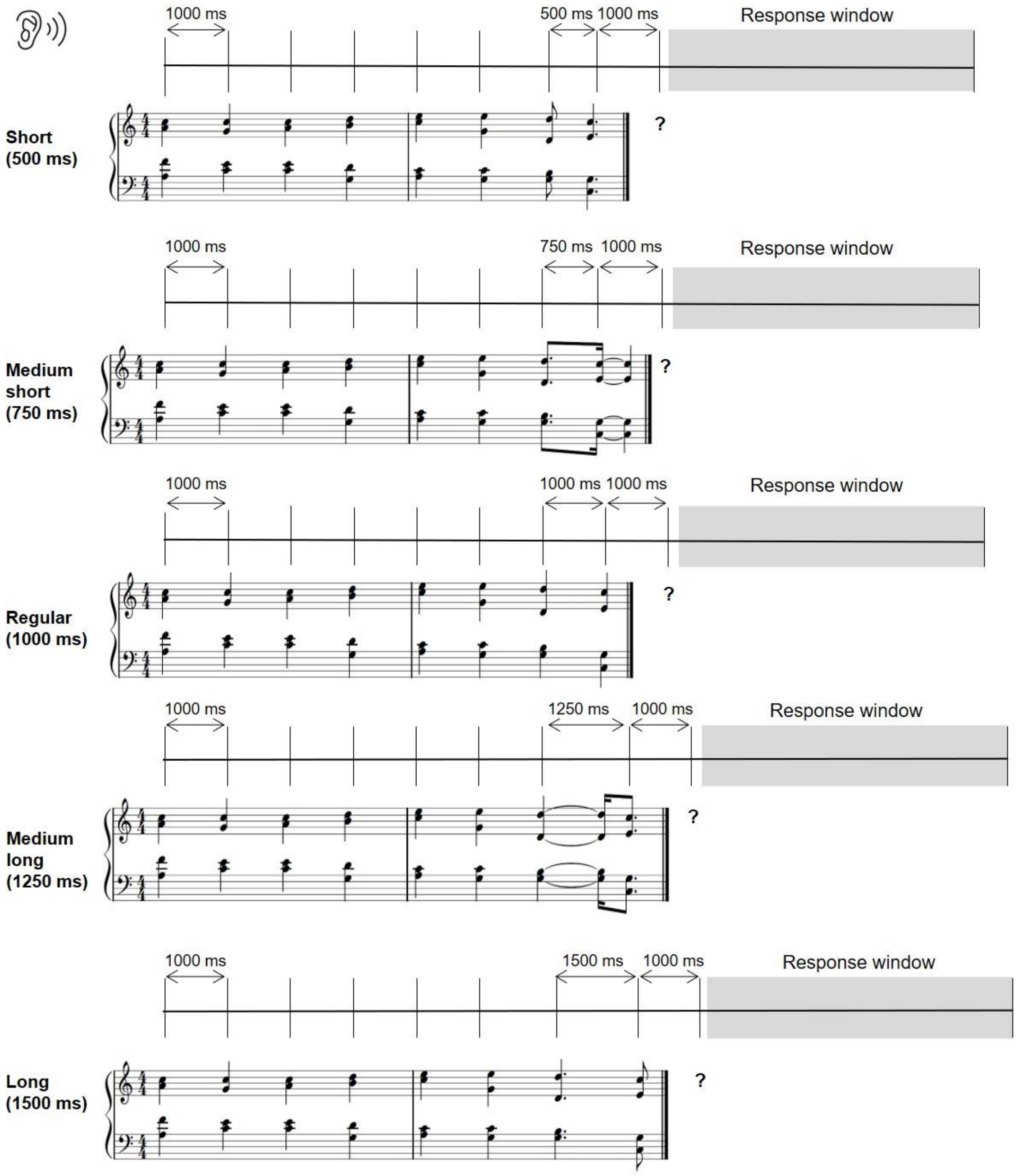


Figure 4.2 – Illustration of the five interstimulus intervals in the music task: short (500 ms), medium-short (750 ms), regular (1000 ms), medium-long (1250 ms) and long (1500 ms). The response window (100–5000 ms after question onset) is shaded in grey.

Additional materials

See section 3.2.1 Experiment 2: Method (Materials) for details of the Language Experience Questionnaire. See section 2.2.2 Experiment 1: Method (Materials) for more information about the Musical Training Questionnaire.

Procedure

Participants sat by themselves in an isolated booth. Auditory material was presented over loudspeakers, and visual material was presented on a computer screen. There were two tasks: the language task and the music task. The order of the tasks was counterbalanced between participants. For both tasks, there were ten practice trials: two sentences and chord progressions were presented with each of the five interstimulus intervals. For both language and music tasks, there were five blocks of 24 experimental trials. The sentences and chord progressions were presented in random order across their five blocks. There were 120 experimental trials in both the language and music tasks, making 240 experimental trials in total. Participants were able to take breaks after each block.

After each sentence or chord progression, a question mark appeared on the computer screen, indicating that the participants should respond to whether the interstimulus interval between the final two words or chords was “Expected” or “Unexpected”, using the response keys. The assignment of the response keys (expected or unexpected, left or right) was counterbalanced between participants. Immediately after the language task, participants were asked to answer five questions about the sentences to check that they had read them. When participants had completed both tasks, they were asked to fill in the Language Experience Questionnaire and the Musical Training Questionnaire.

Analysis

For details of the analysis and graphing software and statistical correction procedures used, see section 2.2.2 Experiment 1: Method (Analysis). In Experiment 4, only trials with responses given between 100–5000 ms after the onset of the question prompt were considered for analysis. Table 4.3 shows the average number of trials that the analyses for each interstimulus interval included.

Table 4.3 – Mean (*SD*) number of trials included after the inclusion criteria were applied in Experiment 4. The maximum possible number of trials for each end type was 24.

	Mean number of trials left (*SD*)	
Interstimulus interval	Music task	Language task
Short (500 ms)	24.0 (0.0)	23.9 (0.3)
Medium-short (750 ms)	24.0 (0.0)	23.9 (0.3)
Regular (1000 ms)	23.9 (0.5)	23.9 (0.3)
Medium-long (1250 ms)	23.9 (0.3)	23.8 (0.8)
Long (1500 ms)	24.0 (0.0)	23.9 (0.3)

4.2.3. Experiment 4: Results

Response frequency

A 5 x 2 x 2 mixed ANOVA was conducted to investigate whether the frequency of “Expected” responses was affected by the interstimulus interval (5 levels within-subjects: 500 ms, 750 ms, 1000 ms (regular), 1250 ms and 1500 ms), task (2 levels within-subjects: language task and music task) or musical training (2 levels between-subjects: amateur musicians and non-musicians). There were interstimulus interval and task effects, and there was an interaction between interstimulus interval and task (discussed below). There were no main effects or interactions associated with musical training (all *p*s > .05).

Effect of interstimulus interval

When data was collapsed across the two tasks (language task and music task), interstimulus interval affected the frequency of “Expected” responses $F(4, 56) = 53.60$, $p < .001$, $\eta p^2 = 0.80$ (see Figure 4.3). Participants rated the regular 1000 ms interstimulus interval as mostly “Expected” *(M* = 95%, *SD* = 1%). They rated the short 500 ms *(M* = 22%, *SD* = 6%) and long 1500 ms *(M* = 23%, *SD* = 6%) interstimulus intervals as mostly unexpected. They rated the medium-short 750 ms *(M* = 47%, *SD* = 4%) and medium-long 1250 ms *(M* = 50%, *SD* = 6%) interstimulus intervals as “Expected” on roughly half the trials and “Unexpected” on the other half.

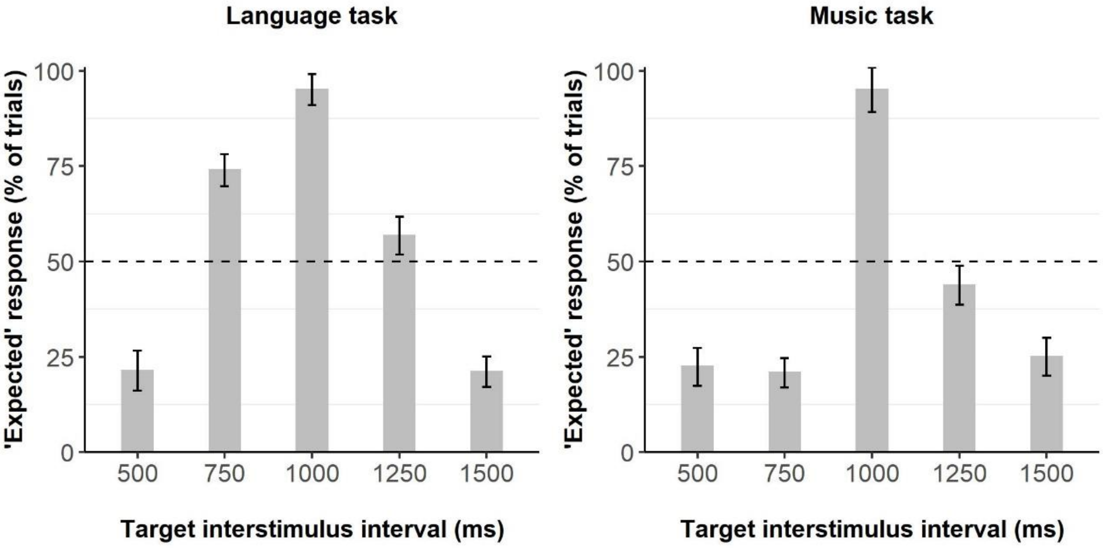


Figure 4.3 – Mean % of trials to which participants responded “Expected” in the language task (left) and the music task (right) for the five interstimulus intervals: short (500 ms), medium-short (750 ms), regular (1000 ms), medium-long (1250 ms) and long (1500 ms). The dashed line illustrates the 50% (chance) level. Within-subject SE bars are included.

There was no difference in the frequency of “Expected” responses between the short 500 ms and long 1500 ms interstimulus intervals (*p* > .05). Similarly, there was no difference between the medium-short 750 ms and medium-long 1250 ms interstimulus intervals (*p* > .05). All other pairwise comparisons were significant after Bonferroni correction (see Table 4.4). The medium-short 750 ms and medium-long 1250 ms were more “Expected” than the short 500 ms and long 100 ms interstimulus intervals. Therefore, there were three levels of expectedness: unexpected (short 500 ms and long 1500 ms), neither expected nor unexpected (medium-short 750 ms and medium-long 1250 ms) and expected (regular 1000 ms).

Table 4.4 – *p*s (after Bonferroni corrections) for pairwise comparisons for the frequency of "Expected" responses between the five interstimulus intervals: short (500 ms), medium-short (750 ms), regular (1000 ms), medium-long (1250 ms) and long (1500 ms).

Interstimulus interval (ms)	500	750	1000	1250	1500
500	–				
750	.001	–			
1000	<.001	<.001	–		
1250	.021	n.s.	<.001	–	
1500	n.s.	.010	<.001	<.001	–

Effect of task

Task also affected the frequency of "Expected" responses $F(1, 14) = 11.61$, $p = .004$, $\eta p^2 = 0.44$ (see Figure 4.3). Overall, the language task received more "Expected" responses *(M* = 54%, *SD* = 10%) than the music task *(M* = 41% *SD* = 10%)*.* This result suggests that participants found the irregular interstimulus intervals less unexpected in the language task compared to the music task.

Interaction between interstimulus interval and task

There was an interaction between interstimulus interval and task for the frequency of "Expected" responses $F(4, 56) = 18.21$, $p < .001$, $\eta p^2 = 0.55$ (see Figure 4.3). *t*-tests were designed to investigate this interaction. Comparisons were made between language and music tasks for each interstimulus interval (500 ms music with 500 ms language etc.). The only significant difference, at the $p < .05$ level, after Bonferroni corrections was for the medium-short 750 ms interstimulus interval $t(15) = 8.52$, $p < .001$, $d = 2.46$. Participants rated the medium-short 750 ms interstimulus interval as more expected in the language task *(M* = 74%, *SD* = 17%) compared to the music task *(M* = 21%, *SD* = 16%).

Reaction time

A second 5 x 2 x 2 mixed ANOVA was carried out to investigate whether reaction time was affected by the interstimulus interval (5 levels within-subjects: short 500 ms, medium-short 750 ms, regular 1000 ms, medium-long 1250 ms and long 1500 ms), task (2 levels within-subjects: language task and music task) or musical training (2 levels between-subjects: amateur musicians and non-musicians). There was an effect of interstimulus interval on reaction time (discussed below). All other effects and interactions were not significant (*p*s >.05).

Effect of interstimulus interval

Interstimulus interval affected reaction time $F(4, 56) = 7.53$, $p < .001$, $\eta p^2 = 0.35$ (see Figure 4.4). The reaction time to the regular 1000 ms interstimulus interval was not significantly different to any of the irregular interstimulus intervals (all *p*s > .05*).* The medium-short 750 ms interstimulus interval was responded to slower *(M* = 670 ms, *SD* = 44 ms) than the three other irregular interstimulus intervals: short 500 ms *(M* = 517 ms, *SD* = 56 ms, *p* = .007), medium-long 1250 ms *(M* = 586 ms, *SD* = 32 ms, *p* = .012) and long 1500 ms *(M* = 543 ms, *SD* = 42 ms, *p* = .004). This result suggests that the medium-short 750 ms interstimulus interval may have been difficult to classify as “Expected” or “Unexpected”. No other pairwise comparisons were significant (all other *p*s > .05).

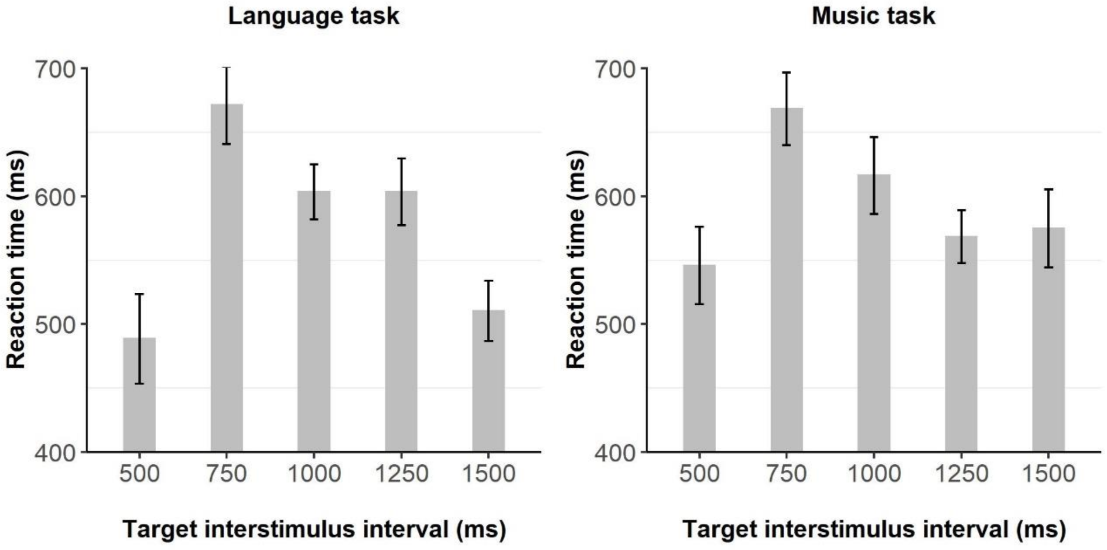


Figure 4.4 – Mean reaction time (ms) in the language task (left) and the music task (right) for the five interstimulus intervals: short (500 ms), medium-short (750 ms), regular (1000 ms), medium-long (1250 ms) and long (1500 ms). Within-subject SE bars are included.

4.2.4. Experiment 4: Discussion

The main aim of Experiment 4 was to identify metric patterns that people reliably find “Expected” and “Unexpected” in both language and music tasks so that, in future studies, ERP correlates of these meter violations can be compared between domains. The results suggest that this aim was achieved.

First, the regular interstimulus interval was reliably rated as “Expected”. It was essential to test that regular interstimulus intervals were rated as “Expected”, to be confident that it was suitable to measure the effects of meter violations against them in future experiments.

Second, the results identify meter violations that were reliably rated as “Unexpected” in both language and music domains and those that were not. The short 500 ms and long 1500 ms interstimulus intervals were rated consistently as “Unexpected” for both the language and music tasks. There were no significant differences in participants’ ratings between tasks for these interstimulus intervals, suggesting that participants found them equally “Unexpected” in both domains.

Additionally, reaction time was not significantly different between the language task and music task, suggesting that task effort was similar for both tasks (Warrick et al., 1965). These findings suggest that it is appropriate for future experiments to compare neural responses associated with these short 500 ms and long 1500 ms interstimulus interval violations between language and music domains.

Third, the results also identified interstimulus intervals that were not reliably rated as "Unexpected" between language and music domains. When data were collapsed across tasks, responses to medium-short 750 ms and medium-long 1250 ms interstimulus intervals were roughly "Expected" on half of the trials, and "Unexpected" on the other half. Furthermore, an interaction between interstimulus interval and task revealed that participants rated the medium-short 750 ms interstimulus interval as more "Expected" in the language task than in the music task, suggesting that participants found it more difficult to distinguish from a regular interstimulus interval in the language task. The medium-short 750 ms interstimulus interval was also responded to more slowly than the three other irregular intervals when ratings were collapsed across tasks. Together, these findings suggest that the medium-short 750 ms interstimulus interval was not perceived as a reliably "Unexpected" interstimulus interval in both language and music domains. Therefore, it is not a suitable candidate for future studies aiming to compare ERP effects associated with meter violations in language and music domains. These results highlight the importance of testing participants' ratings of expectancies before, or while, attempting to investigate their neural correlates. Understanding participants' perceptions of the expectancy of stimuli will facilitate the selection of suitable stimuli and will lead to better-informed interpretations of neural correlates.

Musical training did not have effects on either frequency of "Expected" responses or on reaction time for either the language task or the music task. This result provided support for comparing neural responses between amateur musicians and non-musicians in the following experiments presented in this thesis.

Overall, Experiment 4 successfully identified irregular interstimulus intervals that can be used to compare ERP effects associated with meter violations between language and music domains. The short 500 ms and long 1500 ms interstimulus intervals are used for this purpose in Experiment 5.

4.3. Experiment 5: ERP correlates of meter violations in language and music domains

4.3.1. Experiment 5: Rationale and aims

Experiment 5 was designed to investigate whether there are ERP effects associated with meter violations, and crucially, whether they are comparable between language and music domains. In Experiment 5, short (500 ms) and long (1500 ms) interstimulus intervals were both included as irregular interstimulus intervals. The inclusion of both short and long directions of irregular interstimulus interval allowed the identification of ERP components associated with the meter violations. These were identified if similar effects were found for the short 500 ms and long 1500 ms intervals, compared to the regular 1000 ms interstimulus interval. If just one direction were used, expectation ERP effects would not be distinguishable from the effect of the presentation time. At the same time, with the current experimental design, effects due to time of presentation can be identified as they will show as linear short-regular-long effects.

As discussed in detail in Chapter 1 (section 1.3.3), meter is a crucial aspect of both the language domain (Magne et al., 2016; Zhang & Zhang, 2019) and the music domain (Besson et al., 1997; Fitzroy & Sanders, 2015; Neuhaus & Knösche, 2008; Raij et al., 1997). Meter violations are often associated with N1 ERP effects in both domains, but P2 effects have only been reported in the music domain (Neuhaus & Knösche, 2008; Pereira et al., 2014; Raij et al., 1997). It is hypothesised that this difference could be due to differences in stimulus designs in meter violations between the language domain and music domains – no previous studies seem to have investigated interstimulus interval violations in the language domain. It is possible that the processing of regular beat meter violations shares neurocognitive substrates between language and music domains. In

Experiment 5, a P2 effect for meter violations is investigated in both language and music domains.

The task paradigm in Experiment 5 was the same as in Experiment 3. The expectancy violations were task-irrelevant. For the language task, participants were asked whether the sentences were “Green?” or “Blue?” and for the music task, they were asked whether the chord progression was played by the “Piano?” or “Organ?”. This task paradigm was chosen to avoid the potential occurrence of additional task-related components (as were found when the expectancy violations were task-relevant in Experiment 2).

Additionally, Experiment 5 investigated within- and cross-domain effects of musical expertise on meter violation effects. Previous research suggests that both language expertise (Elmer et al., 2010; Kalender et al., 2013) and music expertise (Vuust et al., 2005; Zhao et al., 2017) improve meter processing. Furthermore, previous studies report that musical expertise is associated with increased N1 and P2 effects for meter violations (Habibi et al., 2014; Jongsma et al., 2005; Kober et al., 2015). In these previous studies, professional musicians were compared with non-musicians. In the current experiment, amateur musicians are compared with non-musicians. Cross-domain expertise effects could indicate shared neural networks between language and music domains (Patel, 2011).

4.3.2. Experiment 5: Method

Ethics

The University Teaching and Research Ethics Committee at The University of St Andrews approved this study, code PS13256 (see Appendix P).

Participants

Twenty university students volunteered to participate in this study (5 male, Age: *M* = 21 years, *SD* = 2 years, range = 18–27 years). Descriptive statistics of participants’

English language and musical training experience are in Table 4.5. Each testing session took 120 minutes, and participants were reimbursed £10. People could not sign up to participate in the study if they had participated in Experiment 4.

Table 4.5 – Descriptive statistics of participants' English language and musical training experience, Experiment 5.

Experience	Measure	*M* (*SD*)	Range
English language	Native English speaking	Native: 15, Non-native: 5	
	Onset (age in years)	1.2 (2.5)	0–8
	English ability (rating 1–10)	9.7 (0.7)	8–10
	% English spoken in the past year	83.5 (18.4)	48–100
Musical training	Number of years practising	6.9 (4.6)	0–15
	Practice time (hours)	1961 (2780)	0–11960
	Overall ability rating (1–5)	3.0 (1.0)	1–5
	Listening per week (hours)	20.9 (21.9)	1–84

Apparatus

For details of the computer set up and response keys, see section 2.2.2 Experiment 1: Method (Apparatus) For details of the EEG system, see section 3.2.1 Experiment 2: Method (Apparatus).

Materials

Sentences

There were 144 sentence presentations in the language task. The 24 sentence stems from Experiments 2, 3 and 4 were used, adapted from Block and Baldwin (2010). As in Experiment 4, only the standard word end type was used. The 24 sentence stems were presented three times, with different interstimulus intervals in between the penultimate word and the final word. The interstimulus interval was either short (500 ms), regular (1000 ms) or long (1500 ms). See Figure 4.5 for an illustration of the three interstimulus intervals in the language task. These 72 sentences were presented twice:

once in green and once in blue, making a total of 144 sentence presentations. For each sentence (presented in both green and blue) participants were asked either “Green?” or “Blue”, to which they answered “Yes” or “No”.

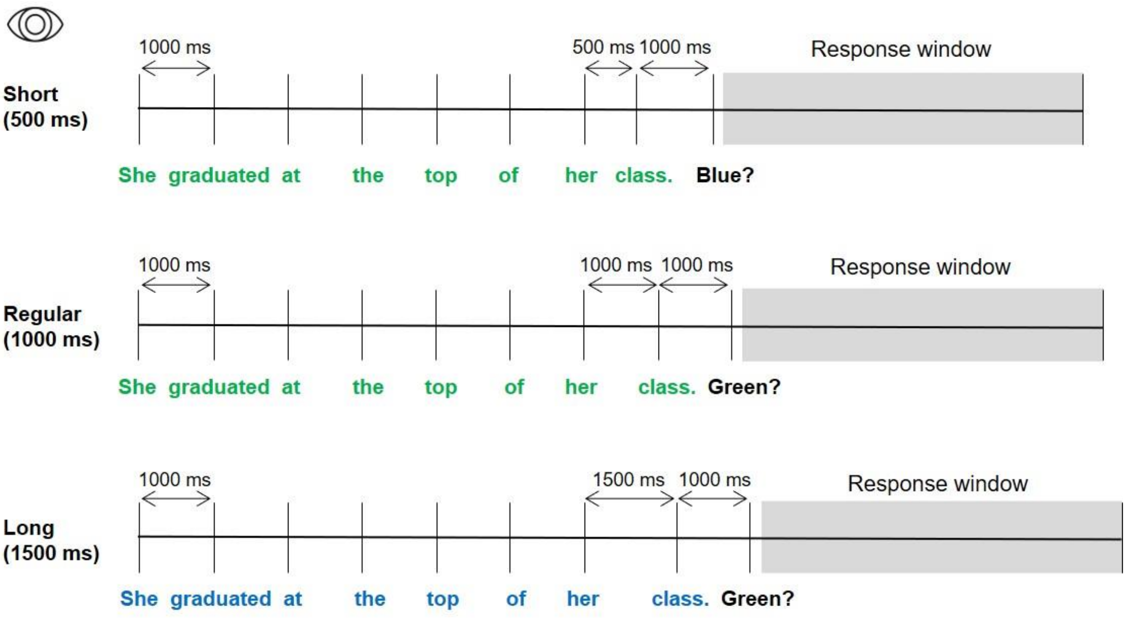


Figure 4.5 – Illustration of the short (500 ms), regular (1000 ms) and long (1500 ms) interstimulus intervals in the language task. The response window (100–5000 ms after question onset) is marked with a shaded area. In these examples, the correct answers (from top to bottom) are: “No”, “Yes”, “No”.

Chord progressions

There were 144 chord progressions in the music task. The 24 chord progression stems from Experiments 2, 3 and 4 were used, adapted from Bach-Gesellschaft (1892). As in Experiment 4, only the tonic (I) end type was presented. These 24 chord progressions were presented three times, with different interstimulus intervals between the penultimate and final chord. The interstimulus interval was either short (500 ms), regular (1000 ms) or long (1500 ms). See Figure 4.6 for an illustration of the three interstimulus intervals in the music task. The 72 chord progressions were presented twice: once played by the piano and once played by the organ. For each chord progression (played by both piano and organ), participants were asked “Piano?” or “Organ?”, to which they answered “Yes” or “No”.

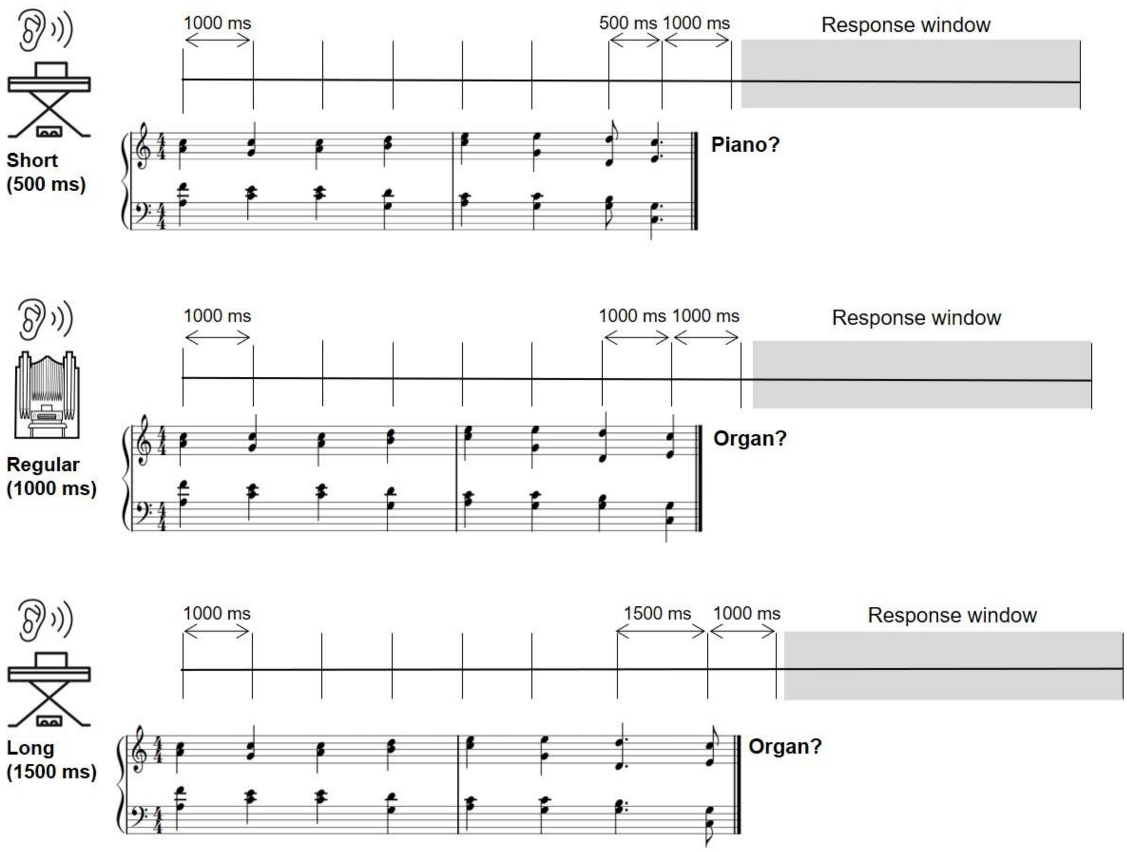



Figure 4.6 – Illustration of the short (500 ms), regular (1000 ms) and long (1500 ms) interstimulus intervals in the music task. The response window (100–5000 ms after question onset) is shaded in grey. In these examples, the correct answers from top to bottom are: “Yes”, “Yes”, “No”.

Additional materials

See section 3.2.1 Experiment 2: Method (Materials) for details of the questions that participants were asked to check they were reading the sentences and the Language Experience Questionnaire. See section 2.2.2 Experiment 1: Method (Materials) for details of the Musical Training Questionnaire.

Procedure

After participants read the information sheet and signed the consent form, they were prepared for EEG recording. Participants sat in an electrically shielded booth. Auditory material was presented over loudspeakers, and visual material was presented on a computer screen. The order of the two tasks (the language task and the music task) was counterbalanced between participants.

For both the language task and the music task, there were nine practice trials. For the practice trials, three sentences and chord progressions were presented with each of the three end interstimulus intervals: short 500 ms, regular 1000 ms and long 1500 ms. 1000 ms after the final word or chord onset, participants were asked whether the sentence was “Blue?” or “Green?” and whether the chord progression was played by “Piano?” or “Organ?”. They responded “Yes” or “No“ using the response keys. Participants chose whether they answered “Yes” with the right hand and “No” with the left (*N* = 16), or vice versa (*N* = 4).

In each task, there were six blocks of 24 experimental trials (288 experimental trials in total). The sentences and chord progressions were presented in random order across their six blocks. Participants were able to take breaks after each block. Immediately after the language task, participants answered the questions about the sentences to check that they had read them. After both tasks were completed, participants were asked to fill in the Language Experience Questionnaire and the Musical Training Questionnaire.

Analysis

For details of the analysis and graphing software and statistical correction procedures used, see section 2.2.2 Experiment 1: Method (Analysis). For more information about the EEG analysis procedures, see section 3.2.1 Experiment 2: Method (Analysis).

Trial inclusion

For Experiment 5, only trials with correct responses given between 100–5000 ms after the onset of the question prompt were considered for analysis. Incorrect responses were not included in the analysis because the task was quite easy, and so incorrect responses may indicate lapses in attention. Moreover, previous studies report ERP effects associated with error processing (Jentzsch et al., 2014; Picton et al., 2012), which could obscure the current effects of interest. Trials were also excluded with artefact rejection using the moving peak-to-peak window method after ICA correction in ERPLAB. Table 4.6 shows the average number of trials in the analysis after the inclusion criteria were applied. For all interstimulus interval types, there were more than 38 trials in the final dataset. Following Luck's (2005) recommendation that ERP studies should be designed with between 30–40 trials per condition with 20 participants to have an acceptable signal-to-noise ratio, assuming that between 10–25% of these will be rejected due to artefacts, the current dataset was considered to have an acceptable signal-to-noise ratio for the analysis.

Table 4.6 – Mean (*SD*) number of trials included in the analysis after the inclusion criteria were applied in Experiment 5. The maximum possible number of trials for each end type was 48.

	Mean number of trials left (*SD*)		
Task	Interstimulus interval	Behaviour	Artefact rejection
Music	Short (500 ms)	46.1 (2.4)	44.0 (3.3)
	Regular (1000 ms)	45.3 (2.8)	43.1 (3.5)
	Long (1500 ms)	45.4 (3.0)	42.4 (4.0)
Language	Short (500 ms)	41.6 (0.8)	39.0 (3.1)
	Regular (1000 ms)	41.3 (0.9)	39.4 (2.1)
	Long (1500 ms)	41.2 (1.1)	38.8 (3.0)

Analysis of ERP effects

As for all ERP analyses presented in this thesis, the analysis consisted of two methods: cluster-based permutation statistics and ANOVA (see section 3.2.1 Experiment 2: Method (Analysis) for further details). The cluster-based permutation statistics were run between 0–800 ms relative to the final word or chord's onset, using the Factorial Mass Univariate ERP Toolbox (Fields, 2017; Groppe et al., 2011). However, as early-onset ERP effects such as these tend to be focal, the likelihood of the cluster-based permutation statistics identifying significant clusters associated with the N1 and P2 effects was significantly reduced (Fields & Kuperberg, 2020). Fortunately, the previous literature provides consistent characterisations of N1 and P2 effects for meter violations. Therefore, a-priori selected electrodes and time windows were chosen for the ANOVA analyses. N1 effects were investigated at Fz between 80–140 ms after the final (target) word's onset or chord, as previous reports of the N1 peak are typically frontal, and to fall within this time window and P2 effects were investigated at Pz between 140–260 ms, as metric violation P2 effects tend to be posterior and measured between this time window (Breen et al., 2019; Menceloglu et al., 2020; Raij et al., 1997). As the data showed that there were not well-defined ERP peaks, the mean amplitudes in these time windows were analysed.

For the previous ERP effects presented in this thesis, the scalp topographies are plotted with the mean amplitude difference method, because this is the method used to analyse the ERP effects in the ANOVA analyses. However, the instantaneous amplitude difference scalp topography map at 200 ms (when the effect is strongest, near its peak) appears to provide clearer insights into the spatial distribution of the P2 effect compared to the mean amplitude difference between 140–260 ms. This issue is discussed further in section 4.3.7, Experiment 5: Discussion. All graphs relating to P2 effects presented in the results sections are presented with the 200 ms instantaneous amplitude method. Both methods are presented for the reader's comparison in Appendix J.

Analysis of expertise effects

To retain statistical power to detect ERP effects, the analyses first include all participants. For any meter violation ERP effects that are significant in both language and music tasks, additional ANOVAs were carried out to investigate whether the effects were significantly different between tasks. For all significant ERP effects, a between-subjects factor of musical training (2 groups: amateur musicians and non-musicians) was added to the ANOVA designs to test for moderating effects of musical expertise.

4.3.3. Experiment 5: Behaviour results

On average, participants answered 74% (*SD* = 31%) of the questions about the sentences correct, providing some evidence that they had read the sentences, and were not just looking at the colour of the sentences.

Percentage of errors

A 3 x 2 within-subjects ANOVA was conducted to investigate whether the percentage of errors that participants made was affected by the interstimulus interval (3 levels: short 500 ms, regular 1000 ms and long 1500 ms) or task (2 levels: language task and music task), see Figure 4.7. The task affected the percentage of errors (discussed below), but there was no effect of interstimulus interval and no interaction between interstimulus interval and task (both *p*s > .05).

Effect of task

Task affected the percentage of errors participants made $F(1, 19) = 67.28$, $p < .001$, $\eta p^2 = 0.78$, see Figure 4.7. Participants made more errors in the language task *(M* = 16%, *SD* = 0.36%) compared to the music task *(M* = 5%, *SD* = 1.31%). In addition to the reaction time results reported below, this result appears to be due to a speed-accuracy trade-off. As well as making more errors in the language task compared to the music task, participants also responded more quickly (see below, Figure 4.8).

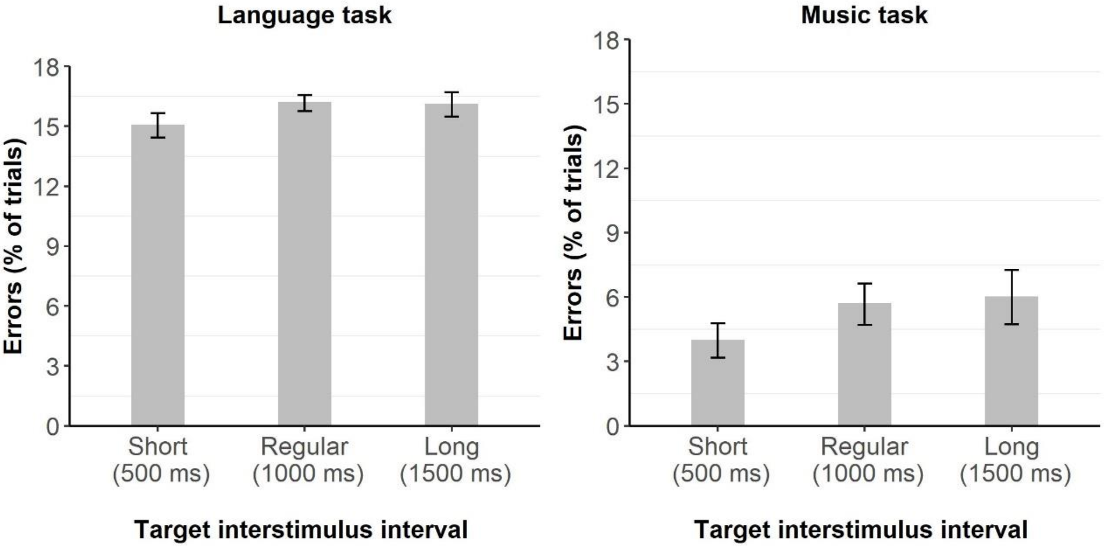


Figure 4.7 – Mean % of errors in the language task (left) and the music task (right) for the three interstimulus intervals: short (500 ms), regular (1000 ms) and long (1500 ms). Within-subject *SE* bars are included.

Reaction time

A 3 x 2 within-subjects ANOVA was carried out to investigate whether reaction time was affected by the interstimulus interval (3 levels: short 500 ms, regular 1000 ms and long 1500 ms) or task (2 levels: language task and music task), see Figure 4.8. There were effects of both interstimulus interval and task on reaction time (discussed below), and no interaction between the two.

Effect of interstimulus interval

Interstimulus interval affected reaction time $F(2, 38) = 5.57$, $p = .008$, $\eta p^2 = 0.23$, see Figure 4.8. Participants responded to the long (1500 ms) interstimulus interval faster (*M* = 715 ms, *SD* = 48 ms) than they did to the regular (1000 ms) interstimulus interval (*M* = 746 ms, *SD* = 48 ms, *p* = .023). No other pairwise comparisons were significant (both *p*s > .05).

Effect of task

Task affected reaction time *F*(1, 19) = 5.85, *p* =.026, ηp^2 = 0.24, see Figure 4.8. Participants responded significantly more quickly in the language task *(M* = 675 ms, *SD* = 149 ms) than the music task *(M* = 789 ms, *SD* = 149 ms). As suggested above, this result appears to be due to a speed-accuracy trade-off. As well as responding more quickly in the language task compared to the music task, participants also made more errors (see above, Figure 4.7).

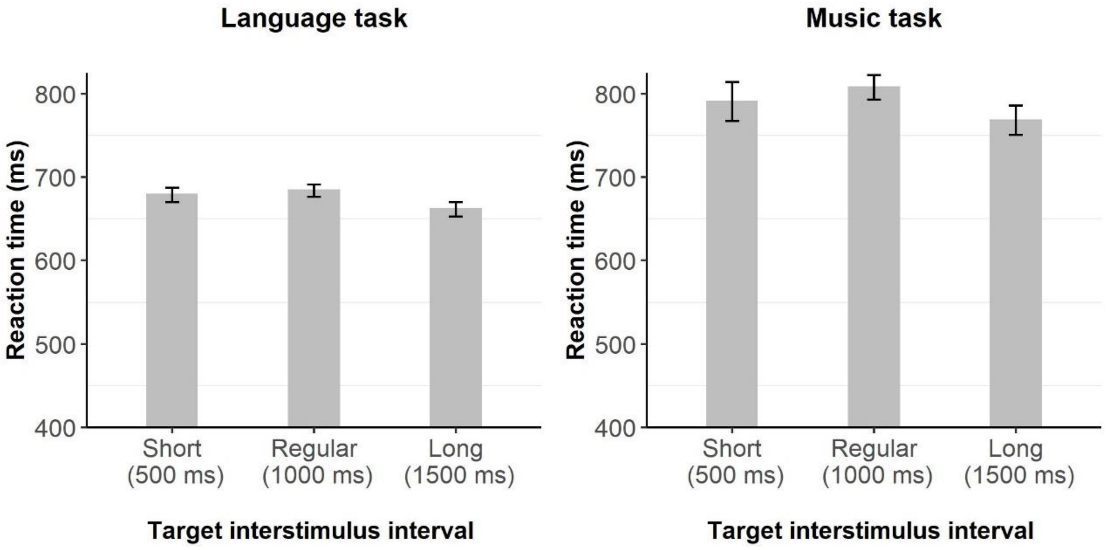


Figure 4.8 – Mean reaction time (ms) in the language task (left) and the music task (right) for the three interstimulus intervals: short (500 ms), regular (1000 ms) and long (1500 ms). Within-subject SE bars are included.

4.3.4. Experiment 5: ERP results

Unforeseen significant clusters were identified by the within-subjects cluster-based permutation statistics in both language and music tasks (see Table 4.7). The tested epoch was between 0–800 ms, and their temporal extents begin at, or very near to, 0 ms. 0 ms is the onset of the final word or chord, and neural responses are unlikely to occur this quickly after stimulus onset (Luck, 2014). Therefore, these clusters appear to represent a pre-stimulus effect.

Table 4.7 – Cluster-based permutation statistics results in language and music tasks, Experiment 5

Task	Mass	*p*	Electrodes	Temporal peak (ms)	Spatial peak	Extent (ms)
Language	4816	<.001	Fp1, AF7, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C3, C5, T7, TP7, CP3, CP1, P1, P3, P5, P7, PO7, PO3, O1, Oz, Pz, Fz, F2, F4, FC4, FC2, C4, TP8, CP4, CP2, P2, P4, P6, P8, PO8, PO4, O2	94	Fz	Start: 0 End: 461
Music	3640	<.001	AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C3, C5, T7, TP7, CP5, CP3, P1, P3, P5, P7, PO7, PO3, O1, Oz, Pz, AF4, Fz, F2, F4, FT8, FC6, FC4, FC2, C4, C6, T8, TP8, CP6, CP4, CP2, P2, P4, P6, P8, PO8, PO4, O2	133	FC1	Start: 16 End: 352

It is currently unclear how to interpret this pre-stimulus effect. The current analysis does not allow for further investigation of this effect, because the epochs were cut at –200 to +1000 ms relative to the onset of the final word or chord and the effect seems to start before this time window.

The temporal characteristics of these pre-stimulus clusters overlap with the a-priori time windows for the N1 and P2 effects. These are quite large clusters so, perhaps unsurprisingly, there are no clear clusters associated with the N1 or P2 time windows. As these pre-stimulus clusters are present in both language and music tasks, the ANOVA analyses aiming to investigate N1 and P2 effects, presented in the following section, were also run with a post-stimulus baseline (0–50 ms). Post-stimulus baselines are seldom recommended for ERP analysis (Luck, 2014). However, in this experiment, the 0–50 ms post-stimulus baseline was used in addition to the –200 to 0 ms pre-stimulus baseline in an attempt to suppress the pre-stimulus effects enough to be able to test the validity of ERP effects associated with the target stimulus (the final chord or word). The following results section refers to both baseline analyses, and full results of the 0 to +50 ms baseline analyses are in Appendix I.

Language meter violations: N1

With the –200 to 0 ms baseline, interstimulus interval appeared to affect mean amplitude at Fz in the N1 time window (80–140 ms after the onset of the final word), $F(2, 38) = 3.76$, $p = .033$, $\eta p^2 = 0.17$. This effect was only significant for the pairwise comparison between regular ($M = –0.14$, $SD = 0.19$) and short 500 ms interstimulus intervals ($M = –0.87$, $SD = 0.17$), $p = .012$, see Figure 4.9. It was not significant for the comparison between regular 1000 ms and long 1500 ms interstimulus intervals ($M = –0.74$, $SD = 0.32$, $p = .316$). There was no statistical difference between short 500 ms and long 1500 ms mean amplitudes ($p = 1.00$). Furthermore, with the 0 to +50 ms baseline, there was no N1 effect $F(2, 38) = 1.21$, $p = .310$, $\eta p^2 = 0.06$, see Appendix I for further details. Overall, the results do not provide clear evidence for an N1 meter violation effect in the language domain.

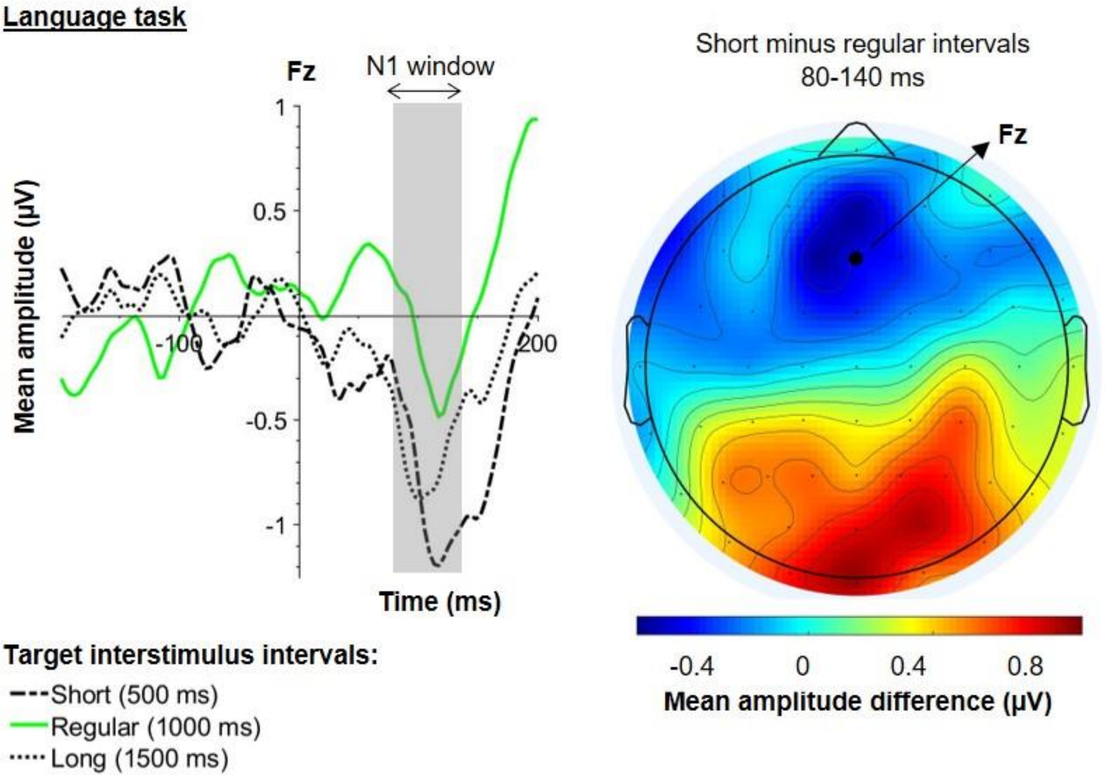


Figure 4.9 – N1 language meter violation graphs (-200 to 0 ms baseline). Left: Mean amplitude (μV) at Fz between –200 to +200 ms for the three interstimulus intervals: short (500 ms), regular (1000 ms) and long (1500 ms). Right: Scalp topography map of the mean amplitude difference (μV) between 80–140 ms after the final (target) word's onset.

Language meter violations: P2

Interstimulus interval manipulations affected the mean amplitude at Pz in the P2 time window (140–160 ms after the onset of the final word), $F(2, 38) = 10.01$, $p < .001$, $\eta p^2 = 0.35$, see Figure 4.10. Mean amplitude was more positive for both the irregular interstimulus intervals, short 500 ms ($M = 1.37$ µV, $SD = 0.40$ µV, $p = .006$) and long 1500 ms ($M = 1.50$ µV, $SD = 0.32$ µV, $p = .001$), compared to the regular–1000 ms interstimulus interval ($M = 0.26$ µV, $SD = 0.25$ µV). There was no amplitude difference between the short 500 ms and long 1500 ms interstimulus intervals ($p = 1.00$). These P2 effects remained with the 0 to +50 ms baseline ($F(2, 38) = 5.65$, $p = .007$, $\eta p^2 = 0.23$, see Appendix I for further details). Therefore, the evidence suggests that there is a P2 meter expectancy violation effect in the language domain.

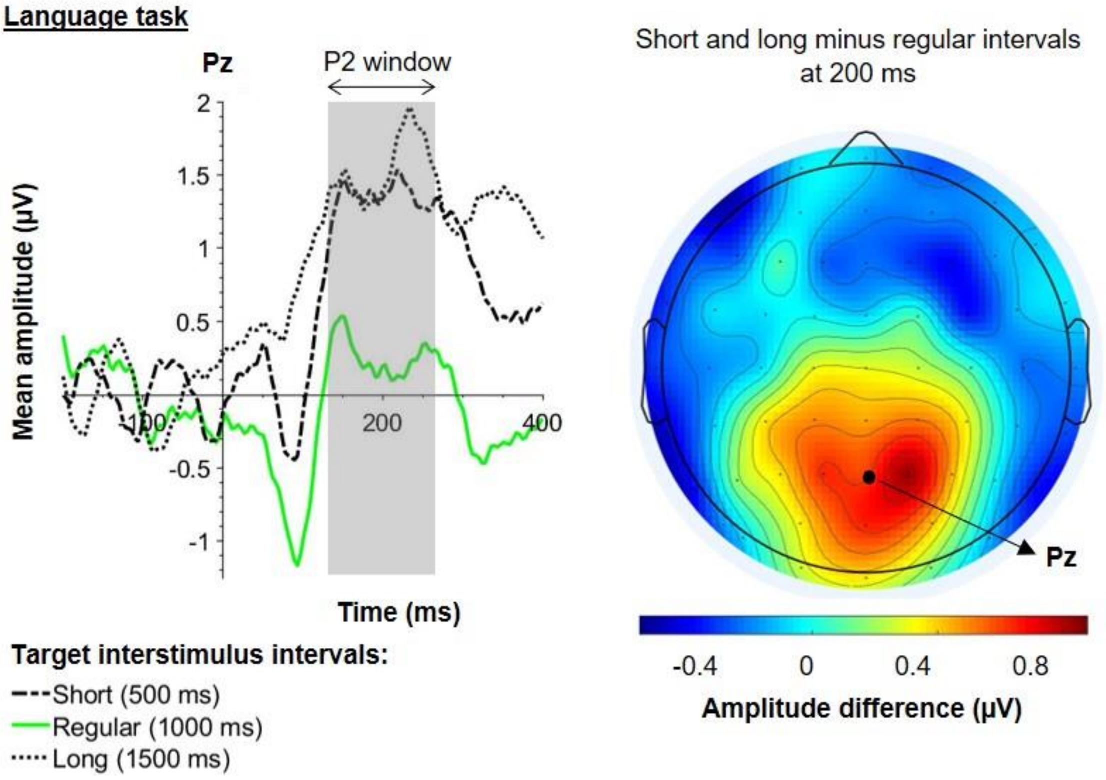


Figure 4.10 – P2 language meter violation graphs (-200 to 0 ms baseline). Left: Mean amplitude (µV) at Pz between –200 to +400 ms for the three interstimulus intervals: short (500 ms), regular (1000 ms) and long (1500 ms). Right: Scalp topography map of the mean amplitude difference (µV) at 200 ms after the final (target) word’s onset.

Music meter violations: N1

Interstimulus interval manipulations affected the mean amplitude at Fz in the N1 time window (80–140 ms after the onset of the final chord), *F*(2, 38) = 7.45, *p* = .002, ηp^2 = 0.28, see Figure 4.11. The mean amplitude was more negative for both the short 500 ms interstimulus interval (*M* = –0.42 µV, *SD* = 0.30 µV, *p* =.018) and the long 1500 ms interstimulus interval (*M* = –0.57 µV, *SD* = 0.23 µV, *p* = .009) compared to the regular 1000 ms interstimulus interval (*M* = 0.35 µV, *SD* = 0.18 µV). There was no amplitude difference between the short 500 and long 1500 ms interstimulus intervals (*p* = 1.00). These results were also significant with the 0 to +50 ms baseline (*F*(2, 38) = 6.63, *p* = .003, ηp^2 = 0.26, see Appendix I for further details). Thus, there was an N1 amplitude effect for meter expectancy violations in the music domain.

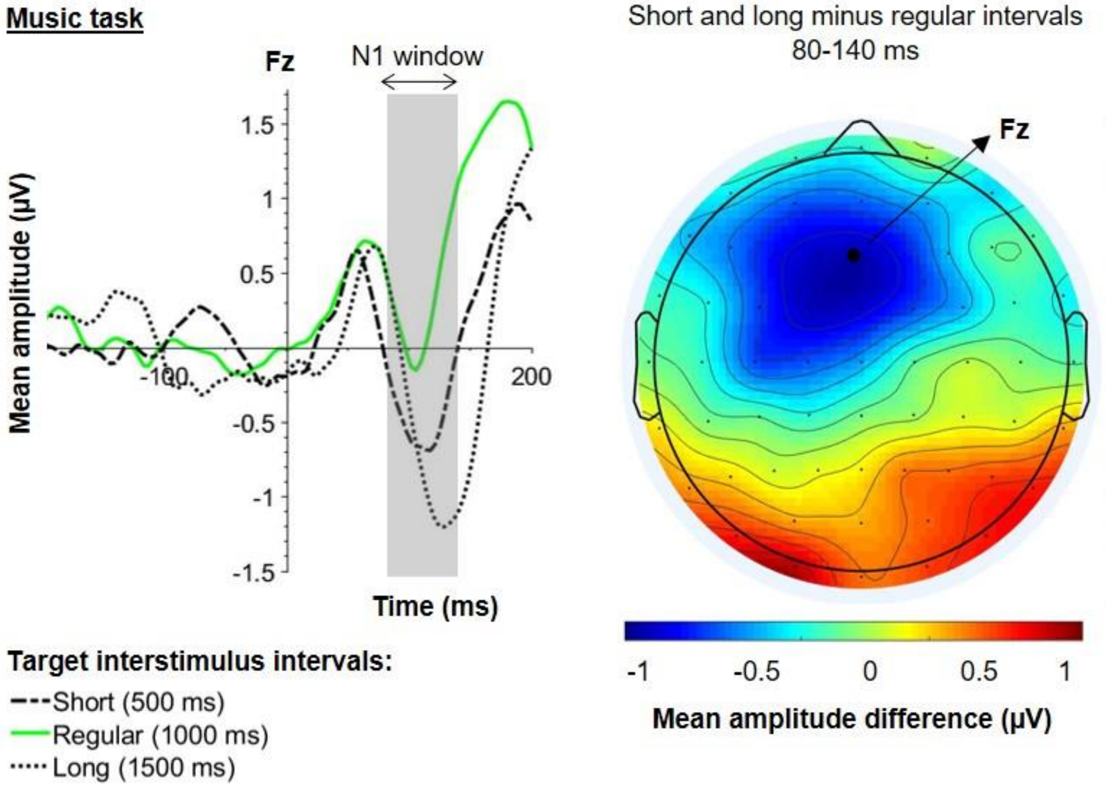


Figure 4.11 – N1 music meter violation graphs (-200 to 0 ms baseline). Left: Mean amplitude (µV) at Fz between –200 to +200 ms for the three interstimulus intervals: short (500 ms), regular (1000 ms) and long (1500 ms). Right: Scalp topography map of the mean amplitude difference (µV) between 80–140 ms after the final (target) chord's onset.

Music meter violations: P2

Interstimulus interval manipulations affected the mean amplitude at Pz in the P2 time window (140–260 ms), *F*(2, 38) = 8.47, *p* = .001, ηp^2 = 0.31, see Figure 4.12. Pairwise comparisons showed an increase in positivity for the short 500 ms interstimulus interval (*M* = 1.20 µV, *SD* = 0.25 µV, *p* = .002) compared to the regular 1000 ms interstimulus interval (*M* = 0.12 µV, *SD* = 0.19 µV). There was also a trend for an increase in positivity for the long 1500 ms interstimulus interval (*M* = 0.85 µV, *SD* = 0.26 µV) compared to the regular–1000 ms interstimulus interval, *p* = .052. There was no difference between short 500 and long 1500 interstimulus intervals (*p* = .641). With the 0 to +50 ms baseline (*F*(2, 38) = 6.78, *p* = .003, ηp^2 = 0.26), these effects and trends remained (see Appendix I for details). Overall, the evidence suggests that there is a P2 effect for meter violations in the music domain.

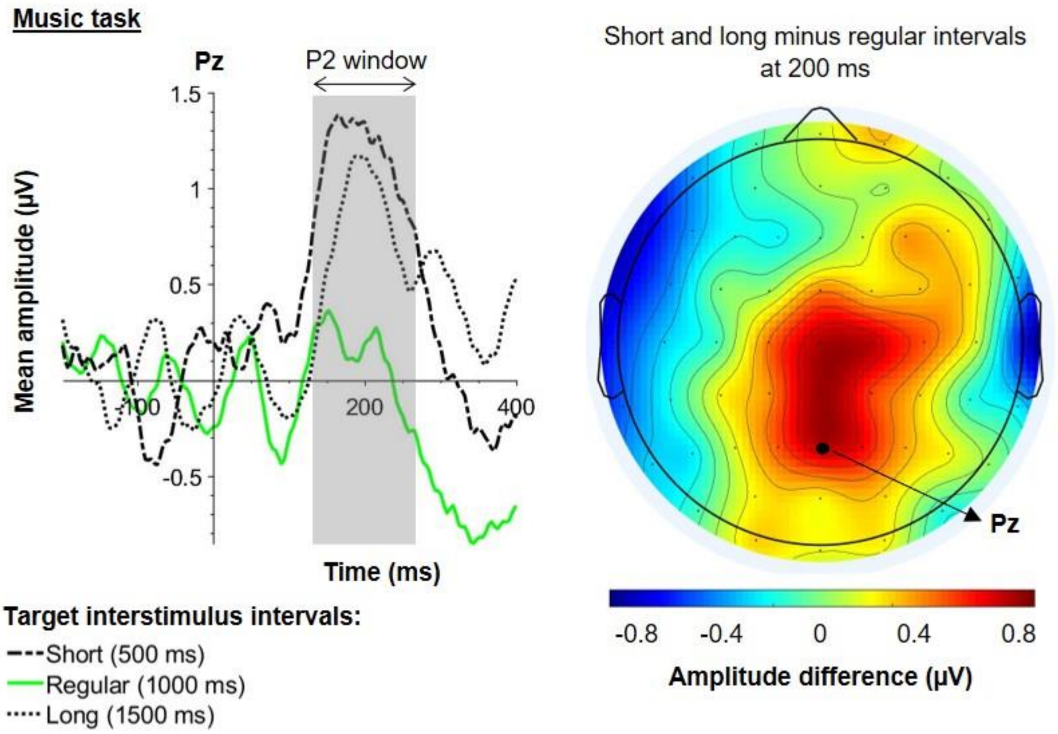


Figure 4.12 – P2 music meter violation graphs (-200 to 0 ms baseline). Left: Mean amplitude (µV) at Pz between –200 to +400 ms for the three interstimulus intervals: short (500 ms), regular (1000 ms) and long (1500 ms). Right: Scalp topography map of the mean amplitude difference (µV) at 200 ms after the final (target) chord’s onset.

4.3.5. Experiment 5: Test of domain-general ERP effects

To aid the interpretation of P2 meter violation effects between language and music domains, additional ANOVAs (3 x 2; interstimulus interval x task) were carried out, for both the –200 to 0 ms and 0 to +50 ms baselines. If there were any interactions between task and interstimulus interval, the case for a domain-general P2 effect would be weakened. Figure 4.13 shows the scalp topography maps of the P2 effects in question, for visual comparison.

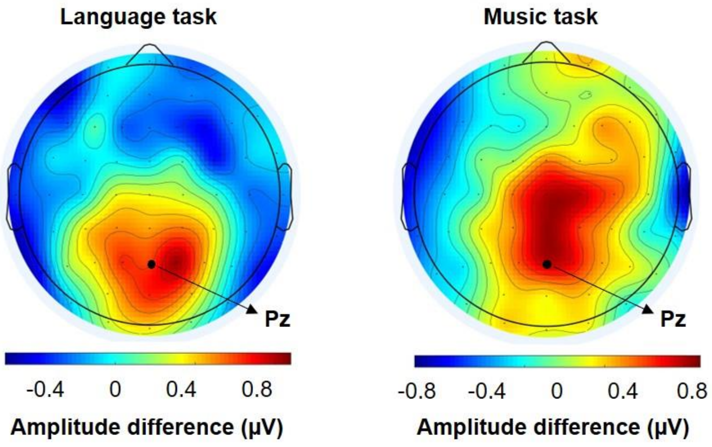


Figure 4.13 – P2 scalp topography maps from the language task (left) and music task (right) in Experiment 5 (-200 to 0 ms baselines). These maps show the amplitude difference for the irregular interstimulus intervals (average of short and long) minus that for the regular interstimulus intervals at 200 ms. These scalp maps are presented separately in the results section above and are presented side by side here to aid comparison. Please note: these scalp maps are on different scales.

The ANOVAs showed that there was an effect of the interstimulus interval at Pz (which showed the P2 effect, as would be predicted, see Appendix I for details). Crucially, there was no effect of task and no interaction between interstimulus interval and task for either the –200 to 0 ms baseline, $F(2, 38) = 1.20$, $p = .311$, $\eta p^2 = 0.06$, or the 0 to +50 ms baseline ($p > .05$), see Appendix I. The lack of interactions between interstimulus interval and task supports (or, at least, does not contradict) the suggestion that the P2 effect appears to be a comparable ERP correlate of meter expectancy violations in both language and music domains.

4.3.6. Experiment 5: Expertise effects

To investigate whether musical training affected within- or cross-domain meter expectation ERP effects, mixed ANOVAs were conducted. A between-subjects factor with two levels (amateur musicians, *N* = 9, and non-musicians, *N* = 11) was added to the ANOVA designs for the significant N1 (music domain) and P2 (language and music domains) ERP effects. As in all experiments in the current thesis, the cut-off was > 1000 hours of practice time for participants to be categorised as amateur musicians. See Table 4.8 for descriptive statistics of each musical training group and Appendix F for the frequency of accumulated practice times.

Table 4.8 – Descriptive statistics of musical training experience for amateur musicians and non-musicians, Experiment 5.

Measure	Amateur musicians (*N* = 9)		Non-musicians (*N* = 11)	
	M (*SD*)	Range	*M* (*SD*)	Range
Number of years practising	10.7 (2.9)	7–15	4.0 (2.8)	1–8
Practice time (hours)	2894 (3246)	1560–11960	385 (353)	26–885
Overall ability rating (1–5)	3.7 (0.9)	2–5	2.5 (0.8)	1–4
Listening per week (hours)	19.8 (17.7)	1–56	21.8 (25.7)	1–84

To investigate whether musical training affected behaviour responses, two 3 x 2 x 2 mixed ANOVAs were carried out to investigate whether i) the percentage of errors participants made on the task or ii) reaction time were affected by the interstimulus interval (3 levels within-subjects: short 500 ms, regular 1000 ms and long 1500 ms), task (2 levels within-subjects: language task and music task) or musical training (2 levels between-subjects: amateur musicians and non-musicians). There were no interactions between musical training and interstimulus interval or task for either percentage of errors or reaction time (all *p*s > .05).

Similarly, for each ERP effect, the N1 (music domain) and P2 (language and music domains), mixed ANOVAs (3 x 2; interstimulus interval x musical training) were conducted to investigate whether musical training affected the meter expectation ERP effects. For all ERP effects, there were no main effects or interactions associated with musical training (all *p*s > .05).

4.3.7. Experiment 5: Discussion

The main aim of Experiment 5 was to investigate whether meter violations elicit ERP effects in language and music domains and, importantly, whether there is any evidence of domain-general neurocognitive substrates for meter processing. The results show that meter violations elicited an N1 effect in the music domain. Crucially, a P2 effect was elicited by meter violations in both language and music domains. This P2 effect could indicate domain-general neurocognitive substrates for meter processing in language and music domains. In this discussion, the pre-stimulus effect, N1, P2 and expertise effects are discussed before briefly considering the wider academic contributions of these results.

The pre-stimulus effects

The within-subjects cluster-based permutation statistics identified pre-stimulus effects in both language and music tasks. Previous studies have identified pre-attentive responses to meter violations (Bouwer et al., 2014; Silva & Castro, 2019). However, in the current experiment, the effect seems to occur before the meter violation. The cluster-based analysis was conducted between 0–800 ms, and the significant clusters begin very near to 0 ms suggesting that they started before the meter violations.

It is currently unclear how to interpret these pre-stimulus effects. Further research, explicitly designed to investigate these effects, is needed. Analysis should be focused on sequences building meter expectancy, rather than –200 to +1000 ms relative to the target interstimulus interval. This approach is recommended because the pre-stimulus effect could be a contingent negative variation (CNV) effect, which is a slow negative wave that builds with the anticipation of the occurrence of a stimulus (Kononowicz & Penney, 2016).

The temporal characteristics of these clusters overlap with the a-priori time windows for the N1 and P2 effects. As a result, there were no clear clusters associated with N1 or P2 effects. For ANOVA analyses involving interstimulus interval in the following experiments presented in this thesis, the 0 to +50 ms baseline will continue to be used in addition to the –200 to 0 ms pre-stimulus baseline, in attempt to test the validity of post-stimulus ERP effects associated with the intended meter expectancy violations (the onset of the final word or chord).

N1

In the language task, the results showed a main effect of mean amplitude for the interstimulus interval in the a-priori N1 time window (80–140 ms) with the –200 to 0 ms baseline. However, pairwise comparisons showed that this was only significant for the short and not for the long interstimulus intervals. Additionally, there was no significant effect of interstimulus interval in the 0 to +50 ms baseline analysis. As a result, there is not clear or strong evidence for an N1 meter violation effect in the language domain. In later experiments, the short 500 ms and long 1500 ms interstimulus intervals are combined to make an “irregular interstimulus interval” condition. Consequently, it was deemed inappropriate for the later experiments to investigate an N1 effect for meter violations in the language domains, as any significant effects of “irregular interstimulus interval” might be due to the short 500 ms interstimulus interval, and not both short 500 ms and long 1500 ms interstimulus intervals. For these reasons, N1 effects for meter violations in the language domain are not investigated in later experiments. The N1 is thought to be sensitive to the attention given to the interstimulus interval. It is larger when attention is given compared to when it is not (Foldal et al., 2020). Therefore, future studies aiming to investigate the N1 effect for meter violations in the language domain could adopt a task that explicitly draws attention to the interstimulus interval between the final two stimuli, like the task in Experiment 4.

In the music task, a clear N1 effect was elicited by meter violations. This effect was frontally distributed and was significant in the a-priori time window (80–140 ms) at

Fz. There was an increase in negativity for both irregular (short 500 ms and long 1500 ms) interstimulus intervals compared to the regular 1000 ms interstimulus interval. This result is consistent with previous experiments that also found a frontal N1 effect for meter violations in the music domain (Fitzroy & Sanders, 2015; Habibi et al., 2014; Kober et al., 2015). Therefore, the N1 effect is potentially a reliable effect for meter violations in the music domain.

While the current experiment does not provide clear evidence for an N1 effect in the language domain, it also does not provide clear evidence against it. This is particularly the case because the N1 is an early ERP effect, which is more prone to noise and measurement error than later effects (Luck, 2014). Future research should continue to compare N1 meter violation effects between language and music domains.

P2

In both language and music tasks, there was a P2 effect for meter violations. These P2 meter violation effects occurred between 140–260 ms after the final (target) word's onset or chord (the a-priori time window). In both tasks, there was an increase in positive amplitude for both irregular (short 500 ms and long 1500 ms) interstimulus intervals compared to the regular 1000 ms interstimulus interval.

These findings are consistent with previous reports that a P2 effect is elicited for meter violations in the music domain (Neuhaus & Knösche, 2008; Raij et al., 1997). Few previous studies have investigated meter violations in the language domain. None appears to have tested interstimulus intervals, despite an underlying sense of regular meter playing a significant role in language comprehension (Cutler & Otake, 2002; Pitt & Samuel, 1990; Port, 2003). Therefore, the current study provides valuable original insights, suggesting that a P2 effect can occur for meter violations in the language domain.

Additional analyses were carried out to compare P2 effects between language and music domains directly. There were no interactions between the interstimulus interval and task for the P2 effect. The statistical similarity of the P2 effect between domains supports the case for potentially domain-general mechanisms for meter processing between language and music domains. Having said this, the P2 in the language domain had a slightly more posterior scalp topography than the P2 in the music domain. Distinct scalp topographies could indicate different underlying neurocognitive substrates. Comparisons of the topography of the P2 between language and music domains are revisited in section 5.3.7, Experiments 6 and 7: Test of domain-general ERP effects. Overall, the P2 meter violation effect, identified for both language and music domains in the current experiment, is a novel indication of domain-general neurocognitive substrates for meter processing. P2 effects are generally thought of as being attention-related effects (Luck, 2014) and, therefore, their apparent associations with meter processing in both language and music domains are consistent with the Dynamic Attending Theory (Jones & Boltz, 1989), which suggests that meter aids the efficient processing of upcoming events, by directing attention to specific time points – this function might not be domain-dependent.

The scalp topography maps provided clearer insights into the spatial features of the P2 effect when the instantaneous amplitude difference method at 200 ms (when the effect was strongest, near its peak) was used, compared to the mean amplitude difference method (between 140–260 ms). Consequently, the instantaneous amplitude difference method was used for the P2 scalp topography maps in the current experiment, and will also be used for the following experiments presented in this thesis. To repeat, both methods are presented for side-by-side visual comparison in Appendix J. This difference in the clarity of the visualisations suggests that the mean amplitude method, with a time window of 140–260 ms, is perhaps not the most appropriate method for measuring the P2 effect. Future studies could consider shorter a-priori time windows around the peak (e.g., 170–230 ms), which could provide more precise insights into P2 characteristics. Nevertheless, P2 analyses in the subsequent experiments presented in this thesis retain the planned 140–260 a-priori time window for ANOVA analyses, so that their results can be directly compared with those of the current experiment, and because it is strongly recommended to stick to a-priori time windows in ERP analyses to avoid *p*-hacking (Luck & Gaspelin, 2017).

Expertise effects

There were no within- or cross-domain effects of musical training on ERP effects associated with meter violations. This result contrasts with previous studies that find effects of musical training on N1 (Jongsma et al., 2005; Kober et al., 2015) and P2 (Jongsma et al., 2005) amplitude for meter violations in the music domain. These previous studies compare professional level musicians and non-musicians. There may have been no expertise effects in the current experiment because amateur levels of musical training do not lead to cross-domain effects in these meter violation ERP effects. Alternatively, the distinction between amateur musicians and non-musicians in the current studies (> 1000 hours for amateur musicians, < 1000 hours for non-musicians) may not have been separate enough to detect effects. This idea is discussed in more detail in section 6.1.6. Because there were no expertise effects, there was no expertise-based evidence to support the hypothesis that there are shared neural networks between language and music domains (Patel, 2011).

Further contributions

Previous studies have hypothesised that delayed presentation of chords would not be experienced as a violation, as delayed endings are common in music for dramatic effect (Zhang et al., 2019). However, both Experiment 4 and Experiment 5 suggest otherwise. Therefore, the current experiments have notable research implications – both short and long irregular interstimulus intervals can be included in investigations of meter violations. Not only is it possible, but perhaps it should be encouraged, to determine whether any effects are just due to the time of presentation (which would be seen in short-regular-long linear effects) or are due to expectancy violations (in which responses to the short and long interstimulus intervals would be different from the regular interstimulus intervals).

It should be noted that the findings of Experiment 5 do not support the existence of neurocognitive substrates shared only between language and music domains. In fact, because the P2 meter violation effect was found both for visual (language task) and auditory (music task) presentations of meter violations, perhaps it is likely that it might occur for other domains. This approach is a valuable direction for further research. For example, future studies could investigate whether the P2 effect also occurs for interstimulus interval-based meter violations in moving visual stimuli or touch sensations (e.g., vibrations to the participants' hands).

4.4. Conclusion

The main aim of the experiments presented in Chapter 4 was to investigate whether there is evidence of similar ERP effects for meter violations between language and music domains. Experiment 4 found that regular interstimulus intervals are reliably rated as being "Expected", and it successfully identified irregular interstimulus intervals that are reliably rated as being "Unexpected" in both language and music domains. Experiment 5 identified an N1 effect as a meter violation effect in the music domain. There was no clear evidence for an N1 effect for meter violations in the language domain, and this effect is not investigated further in the following experiments presented in this thesis. Crucially, a P2 effect occurred for meter violations in both language and music domains. The P2 effect could indicate domain-general neurocognitive substrates for meter processing in language and music domains.

The ERP effects for meter violations found in Experiment 5 (the N1 in the music domain and P2 in both language and music domains) are revisited in the next chapter, Chapter 5 (in Experiments 6 and 7), when it is tested whether are interactions in these effects when each violation type is presented simultaneously with semantic, grammar and harmonic expectancy violations. If the P2 is replicated for meter violations in both language and music domains, and if it does not interact with other types of expectancy violations, the case for it indicating separable domain-general neurocognitive substrates that are specialised for meter processing could be strengthened.

5. Chapter 5 – Interactions in ERP effects for simultaneous expectancy violations

5.1. Introduction

Chapters 3 and 4 indicated distinct ERP effects elicited by semantic, grammar and harmonic expectancy violations (Experiments 2 and 3) and a potentially domain-general ERP effect elicited by meter violations (Experiment 5). In the current chapter, Chapter 5, it is investigated whether there are interactions in these ERP effects when meter violations are presented simultaneously with semantic, grammar and harmonic expectancy violations.

The approach of investigating interactions in ERP effects for simultaneously presented expectancy violations has been taken in previous studies. These previous studies suggest that an absence of interactions in ERP effects supports the hypothesis that there are separable neurocognitive substrates underpinning the them (Gunter et al., 2000; Wicha et al., 2004). Therefore, interactions between meter violation ERP effects and semantic, grammar or harmony ERP effects could provide evidence that their underpinning neurocognitive substrates are linked. This finding would call into question the existence of domain-specialised neurocognitive substrates for semantic, grammar and harmony processing and domain-general neurocognitive substrates that are specific to meter processing.

In both Experiments 6 and 7, the task was irrelevant to the expectancy violations. Therefore, the task did not bias towards ERP effects based on violation type, and the results can be more directly compared to those of Experiments 3 and 5. In Experiment 6, semantic and grammar expectancy violations were presented simultaneously with meter violations. In Experiment 7, harmonic violations were presented simultaneously with meter violations. As the previous experiments presented in this thesis have found robust ERP effects in within-subjects designs, it was of interest whether these would also be found in between-subjects designs. This was particularly of interest for the meter violations – as, although task order was counterbalanced between participants in Experiment 5, meter violations in one domain may have primed meter violations in the other domain, affecting the neurocognitive substrates that were recruited for their processing.

Additional analyses were conducted, where sample characteristics allowed[9], to investigate the potentially moderating within- or cross-domain expertise effects of English language experience and musical training on significant expectancy violation ERP effects. Cross-domain expertise effects could indicate shared neurocognitive substrates between language and music domains (Patel, 2011).

5.2. Experiment 6: Language domain experiment – simultaneous presentation of meter violations with semantic and grammar expectancy violations

5.2.1. Experiment 6: Rationale and aims

As discussed in more detail in Chapter 1 (section 1.3.4), previous studies investigating links between meter processing and both semantic and grammar processing provide mixed results. One study reported that the N400 for semantic violations was larger when semantic violations were presented with regular meter compared to when they were presented with irregular meter (Li et al., 2019). In contrast, Rothermich et al. (2010) report that the N400 was significantly smaller for a regular meter than an irregular meter. Both authors suggest that their findings might indicate shared neurocognitive substrates for semantic and meter processing. However, other studies find no interactions between meter and semantic expectancy violations in the N400 effect (Magne et al., 2007).

Similarly, there are mixed findings for links between grammar and meter processing. An irregular meter appears to decrease performance on grammar tasks (Chern et al., 2018; Gordon et al., 2015). Some studies report a larger P600 for grammar violations presented with a regular meter than an irregular meter (Canette et al., 2020). Others report the opposite effect – a smaller P600 for grammar violations presented with

---

[9] In Experiment 6, there was not an even split of native (*N* = 14) and non-native (*N* = 6) speakers for analysis of language expertise effects. Consequently, effects of language expertise were not investigated in Experiment 6.

a regular meter compared to an irregular meter (Schmidt-Kassow & Kotz, 2009). Therefore, the current evidence does not provide a clear view of whether there are shared or separable neurocognitive substrates for meter processing and semantic or grammar processing. These previous studies tend not to report ERP effects associated with meter manipulations. Including these in analyses could provide new insights.

5.2.2. Experiment 6: Method

Ethics

The University Teaching and Research Ethics Committee at The University of St Andrews approved this study, code PS13256 (see Appendix P).

Participants

Twenty university students volunteered to participate in this study (4 male, Age: *M* = 21 years, *SD* = 4 years, range = 18–32 years). Descriptive statistics of participants' English language experience and musical training are in Table 5.1. Each testing session took 120 minutes, and participants were reimbursed £10. People could not sign up to participate in the study if they had participated in Experiment 2, 3, 4 or 5.

Table 5.1 – Descriptive statistics of participants' English language and musical training experience, Experiment 6.

Experience	Measure	*M* (*SD*)	Range
English language	Native English speaking	Native: 14, Non-native: 6	
	Onset (age in years)	1.9 (3.5)	0–10
	English ability (rating 1–10)	9.4 (1.5)	5–10
	% English spoken in the past year	74.9 (19.2)	35–100
Musical training	Number of years practising	6.9 (4.7)	0–14
	Practice time (hours)	2224 (2392)	0–9568
	Overall ability rating (1–5)	2.7 (1.2)	1–4
	Listening per week (hours)	12.0 (11.0)	1–50

Apparatus

For details of the computer set up and response keys, see section 2.2.2 Experiment 1: Method (Apparatus) For more information about the EEG system, see section 3.2.1 Experiment 2: Method (Apparatus).

Materials

Sentences

There were 288 sentence presentations in the main experiment. The 24 sentence stems from Experiments 2, 3, 4 and 5 (adapted from sentence stems by Block and Baldwin (2010)) were each presented with three sentence end types (standard word, unrelated semantic violation and grammar expectancy violation) and two types of the interstimulus interval between the final two words (regular 1000 ms and irregular 500/1500 ms). For the irregular interstimulus intervals, half were short (500 ms) and half were long (1500 ms). Therefore, each of the 24 sentence stems was presented six times (see Table 5.2). These 144 sentences were presented twice: once in green and once in blue, making a total of 288 sentence presentations. For each sentence (presented in both green and blue) participants were asked either “Green?” or “Blue?”, to which they answered “Yes” or “No”. For the practice trials, two additional sentence stems were presented with each of the six presentations (see Table 5.2).

Table 5.2 – The six presentation types of the 24 sentence stems.

Sentence end type		Interstimulus interval
Standard word	and	Regular (1000 ms)
Standard word	and	Irregular (500/1500 ms)
Unrelated semantic violation	and	Regular (1000 ms)
Unrelated semantic violation	and	Irregular (500/1500 ms)
Grammar violation	and	Regular (1000 ms)
Grammar violation	and	Irregular (500/1500 ms)

Additional materials

See section 3.2.1 Experiment 2: Method (Materials) for details of the questions participants were asked to check they were reading the sentences and the Language Experience Questionnaire. See section 2.2.2 Experiment 1: Method (Materials) for details of the Musical Training Questionnaire.

Procedure

After participants read the information sheet and signed the consent form, they were prepared for EEG recording. Participants sat in an electrically shielded booth. The visual material was presented on a computer screen. There were 12 practice trials, with two presentations of each of the six presentations (refer to Table 5.2). 1000 ms after the onset of the final word, participants were asked whether the sentence was “Blue?” or “Green?” They responded “Yes” or “No” using the response keys. Participants chose whether they answered “Yes” with the right hand and “No” with the left (*N* = 12), or vice versa (*N* = 8). After the practice trials, the experimental trials began. There were 12 blocks of 24 experimental trials each. The sentences were presented in random order across the 12 blocks. Participants were able to take breaks after each block. After the task was completed, participants answered the questions about the sentences to check that they had read them. They were then asked to fill in the Language Experience Questionnaire and the Musical Training Questionnaire.

Analysis

For details of the analysis and graphing software and statistical correction procedures used, see section 2.2.2 Experiment 1: Method (Analysis). For details of EEG analysis procedures and graphing methods, see section 3.2.1, Experiment 2: Method (Analysis).

Trial inclusion

As in the previous experiments with the implicit tasks, only trials with correct responses given between 100–5000 ms after the onset of the question prompt were considered for analysis. Trials were also excluded with artefact rejection using the moving peak-to-peak window method after ICA correction in ERPLAB. Table 5.3 shows the average number of trials that were included in the analysis after the inclusion criteria were applied. All end types had an average of 44 or more trials included in the final dataset, so based on Luck's (2014) recommendations, the final dataset had an acceptable signal-to-noise ratio.

Table 5.3 – Mean (SD) number of trials included in the analysis after the inclusion criteria were applied in Experiment 6. The maximum possible number of trials for each of the six presentation types = 48.

		Mean number of trials left (*SD*)	
Sentence end type	Interstimulus interval	Behaviour	Artefact rejection
Standard word	Regular (1000 ms)	47.3 (0.9)	44.5 (3.8)
Standard word	Irregular (500/1500 ms)	47.4 (1.0)	45.3 (2.6)
Unrelated semantic violation	Regular (1000 ms)	47.5 (0.8)	45.4 (2.2)
Unrelated semantic violation	Irregular (500/1500 ms)	47.1 (1.2)	44.0 (3.5)
Grammar violation	Regular (1000 ms)	47.6 (1.2)	45.7 (2.4)
Grammar violation	Irregular (500/1500 ms)	47.3 (0.9)	44.7 (3.8)

Analysis of ERP effects

Cluster-based permutation statistics were run between 0–800 ms, relative to target stimulus onset (the onset of the final word), using the Factorial Mass Univariate ERP Toolbox (Fields, 2017; Groppe et al., 2011). The ANOVA analyses parameters used for the ERP effects are the same as in previous chapters, to allow for direct comparison (see Table 5.4). Refer to sections 3.2.1, Experiment 2: Method (Analysis) and 3.3.2, Experiment 3: Method (Analysis) for details of the N400 and P600 analysis choices and section 4.3.2, Experiment 5: Method for justification of the P2 analysis choices.

Table 5.4 – A-priori selected electrodes and time windows for analysis of the P2, N400 and P600 components

ERP component	Electrode	Time (ms)	Violation type
P2	Pz	140–260	Meter
N400	Pz	300–450	Semantic
P600	Cz	500–800	Grammar

To control the pre-stimulus effect associated with interstimulus intervals, discovered in Experiment 5, a post-stimulus (0 to +50 ms) baseline analysis was included in addition to the usual pre-stimulus (–200 to 0 ms) baseline analysis. This analysis was done to test the validity of meter expectancy violation effects that are associated with the onset of the final word (the target interstimulus interval stimulus).

Analysis of expertise effects

As in previous experiments, to retain statistical power to detect ERP effects, the ANOVAs first included all participants. Then, for any significant ERP effects, a between-subjects factor of musical training (2 groups: amateur musicians and non-musicians) was added to the ANOVAs to test for expertise effects.

5.2.3. Experiment 6: Behaviour results

On average, participants answered 74% (*SD* = 27%) of the questions about the sentences correct, providing some evidence that they had read the sentences and were not just looking at the colour of the sentences.

Percentage of errors

A 3 x 2 within-subjects ANOVA was conducted to investigate whether sentence end type (3 levels: standard word, unrelated semantic violation and grammar expectancy violation) or interstimulus interval (2 levels: regular (1000 ms) and irregular (500/1500 ms) affected the percentage of errors that participants made, Figure 5.1. There was a main effect of interstimulus interval, $F(1, 19) = 4.45$, $p = .048$, $\eta p^2 = 0.19$, see Figure 5.1, right panel. There were fewer errors for the regular interstimulus interval *(M* = 0.95%, *SD* = 0.34%) compared to the irregular interstimulus interval *M* = 1.49%, *SD* = 0.40%). There was no effect of sentence end type and no interaction between interstimulus interval and sentence end type (*p*s > .05).

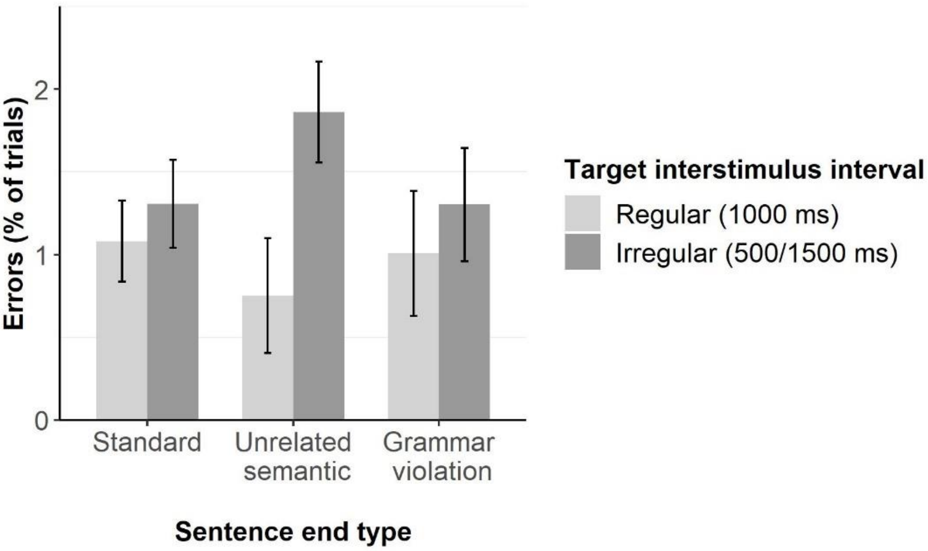


Figure 5.1 – Mean % of errors for sentence end types and interstimulus intervals. Within-subject SE bars are included.

Reaction time

A 3 x 2 within-subjects ANOVA was carried out to investigate whether sentence end type (3 levels: standard word, unrelated semantic violation and grammar expectancy violation) or interstimulus interval (2 levels: regular (1000 ms) and irregular (500/1500 ms) affected reaction time, see Figure 5.2. Sentence end type affected reaction time $F(2, 38) = 11.26$, $p < .001$, $\eta p^2 = 0.37$. Participants responded faster to the standard word *(M*

= 721 ms, *SD* = 45 ms) than both the unrelated semantic violation *(M* = 764 ms, *SD* = 44 ms) and the grammar expectancy violation *(M* = 748 ms, *SD* = 43 ms). There was no difference in reaction time between unrelated semantic and grammar expectancy violations (*p* > .05). There was no effect of interstimulus interval and no interaction between interstimulus interval and sentence end type (both *p*s > .05).

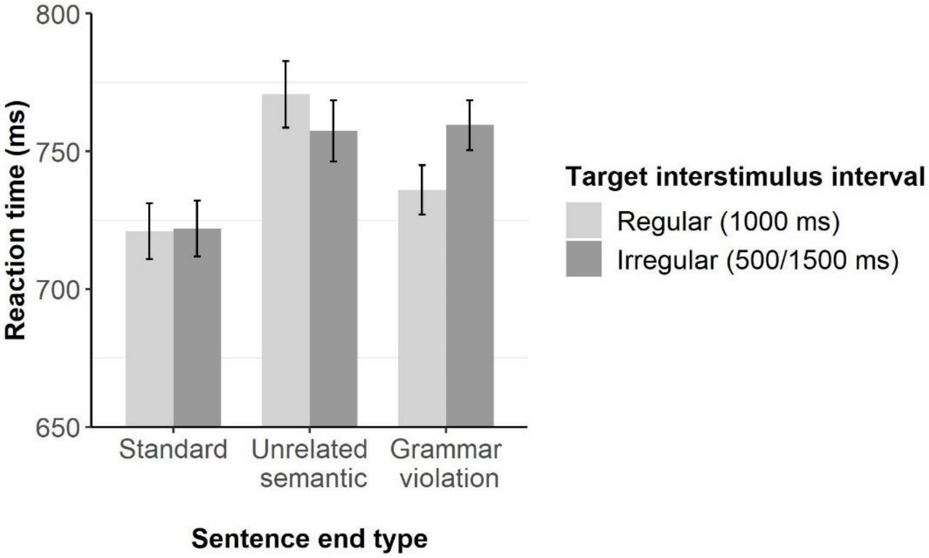


Figure 5.2 – Mean reaction time (ms) for sentence end types and interstimulus intervals. Within-subject SE bars are included.

5.2.4. Experiment 6: ERP results

As in Experiment 5, there were pre-stimulus effects for the effect of interstimulus interval in Experiment 6, which were captured by a cluster in the within-subjects cluster-based permutation statistics (see Table 5.5), carried out between 0–800 ms after the final (target) word's onset. This effect affected the pre-stimulus baseline (–200 to 0 ms) for the interstimulus interval manipulation. Therefore, for all significant ERP analyses were also run with a post-stimulus baseline (0–50 ms), to assess the validity of the effects. Both baseline analyses are referred to in the results section. The results of the 0 to +50 ms baseline analysis are in Appendix K. There were no significant clusters for the sentence end types (for either the standard word and semantic violation cluster analysis, or the standard word and grammar violation analysis) and, crucially, no significant clusters for the test of interactions between sentence end types and interstimulus interval.

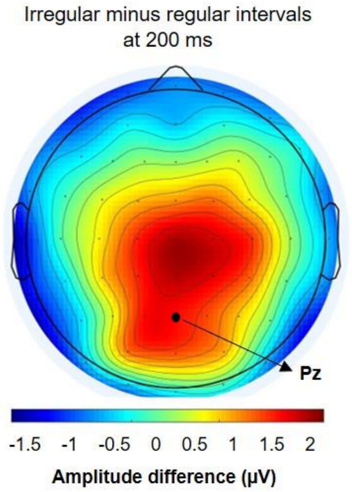
Table 5.5 – Cluster-based permutation statistics results for the effect of the interstimulus interval, data collapsed across sentence word end type.

Mass	*p*	Electrodes	Temporal peak (ms)	Spatial peak	Extent (ms)
4454	<.001	P1, PO3, O1, Oz, Pz, Fp2, AF4, F4, F6, FC4, C4, CP4, CP2, P2, P4, PO4, O2	94	CP2	Start: 0 End: 621

To investigate ERP effects, 2 x 2 within-subjects ANOVAs were conducted separately for semantic and grammar expectancy violations. These ANOVAs were conducted for each ERP effect found previously for the semantic and grammar expectancy violations in Experiments 2, 3 (N400, P600) and the meter violations in Experiment 5 (P2). The aim was to investigate whether sentence end type (2 levels: standard words and unrelated semantic violations or grammar expectancy violations) or interstimulus interval (2 levels: regular 1000 ms and irregular 500/1500 ms) elicited these effects and, crucially, to investigate potential interactions.

P2

Interstimulus interval manipulations affected the mean amplitude at Pz in the P2 time window (140–160 ms), in both analyses: semantic (*F*(1, 19) = 14.03, *p* = .001, ηp^2 = 0.43) and grammar (*F*(1, 19) = 18.10, *p* < .001 ηp^2 = 0.49), see Figure 5.3. For both analyses, mean amplitudes were more positive for the irregular interstimulus intervals (semantic: *M* = 1.40 µV, *SD* = 0.24 µV and grammar: *M* = 1.50 µV, *SD* = 0.26 µV) compared to the regular interstimulus intervals *(M* = 0.53 µV, *SD* = 0.16 µV and *M* = 0.55 µV, *SD* = 0.18 µV, respectively). These effects remained with the 0 to +50 ms baseline (see Appendix K). Crucially, in both analyses, there was no effect of sentence end type (*p*s > .05) and no interaction between meter violations and sentence end type (semantic: *F*(1, 19) = 0.96, *p* = .341, ηp^2 = 0.05, grammar: *F*(1, 19) = 0.27, *p* = .607, ηp^2 = 0.01, see Appendix L for 0 to +50 ms baseline result (*p* > .05)



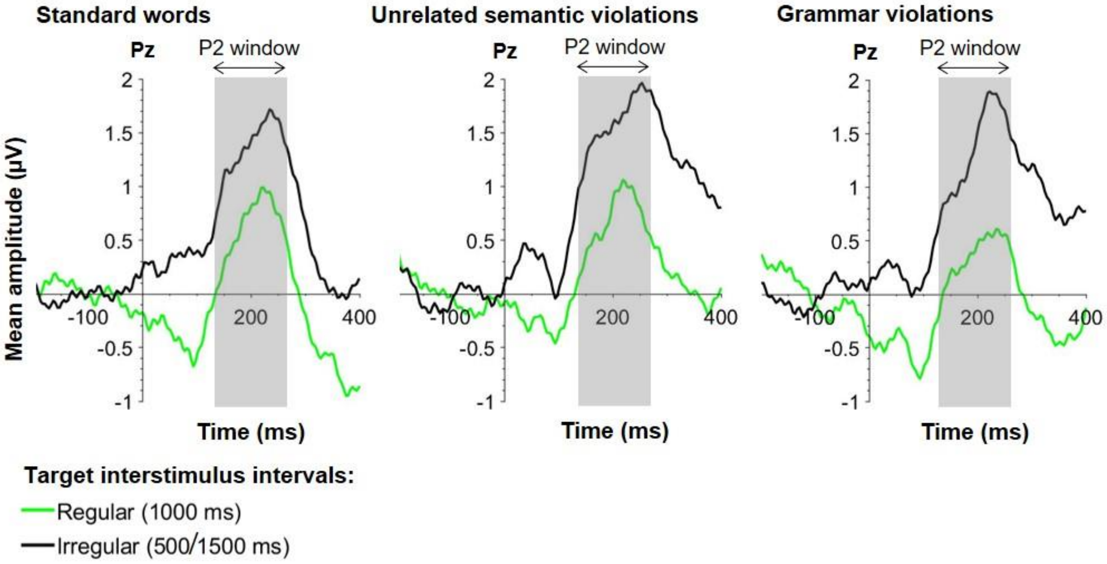


Figure 5.3 – P2 meter violation graphs (–200 to 0 ms baseline). Top right: Scalp topography map of the mean amplitude difference (µV) at 200 ms after the final (target) word's onset. Bottom: ERP graphs at Pz between –200 to +400 ms for the regular and irregular interstimulus intervals, separate for the three sentence end types: standard (left), unrelated semantic violation (middle) and grammar expectancy violation (right).

N400

Sentence end type, with two levels: standard words and unrelated semantic violations, affected the mean amplitude at Pz in the N400 time window (300–450 ms), *F*(1, 19) = 5.92, *p* = .025, ηp^2 = 0.24, see Figure 5.4. There was an increase in negativity for the unrelated semantic violations *(M* = –0.23 µV, *SD* = 0.26 µV) compared to the standard words *(M* = 0.27 µV, *SD* = 0.27 µV). This effect was also found for the 0 to +50 ms baseline analysis, see Appendix K. The within-subjects cluster-based permutation statistics identified a significant main effect cluster (mass = 847, *p* =.027), corresponding to the N400 in centroparietal areas (CP3, P1, P3, PO3, Oz, Pz, C4, CP4, CP2, P2, P4, PO4, O2) between 289–445 ms. No additional clusters were identified for semantic expectancy violations. Importantly, there was no interaction between semantic expectancy violations and meter violations *F*(1, 19) = 0.08, *p* = .785, ηp^2 = 0.004, for either baseline (see Appendix L).

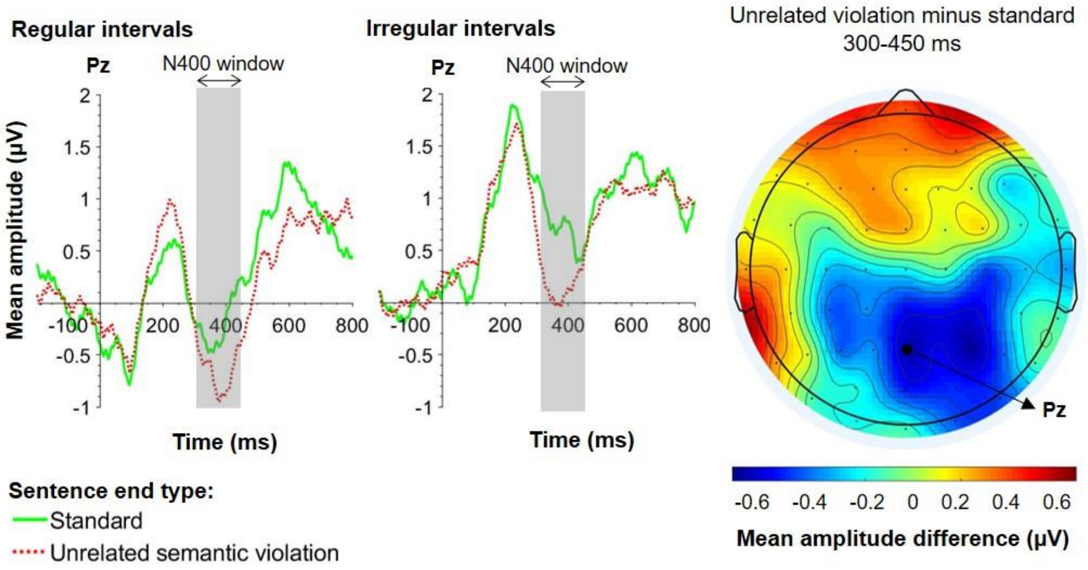


Figure 5.4 – N400 semantic expectancy violation graphs (–200 to 0 ms baseline). Left and middle: Mean amplitude (µV) at Pz between –200 to +800 ms for the two sentence end types: standard word and unrelated semantic violations for regular 1000 ms (left) irregular 500/1500 (middle) interstimulus intervals. Right: Scalp topography map of the mean amplitude difference (µV) between 300–450 ms after the final (target) word's onset.

P600

Sentence end type, with two levels: standard words and grammar violations, affected the mean amplitude at Cz in the P600 time window (500–800 ms after the final (target) word's onset), *F*(1, 19) = 4.63, *p* = .045, ηp^2 = 0.20, see Figure 5.5. The mean amplitude was more positive for grammar expectancy violations *(M* = –0.04 µV, *SD* = 0.17 µV) compared to the standard words *(M* = –0.73 µV, *SD* = 0.30 µV). This effect was also found for the 0 to +50 ms baseline analysis, see Appendix K. The within-subjects cluster-based permutation statistics did not identify any significant clusters associated with the grammar expectancy violations. Critically, there was no interaction between grammar expectancy violations and meter violations, *F*(1, 19) = 0.32, *p* = .580, ηp^2 = 0.02, see Appendix L for 0 to +50 ms baseline interaction result (*p* > .05).

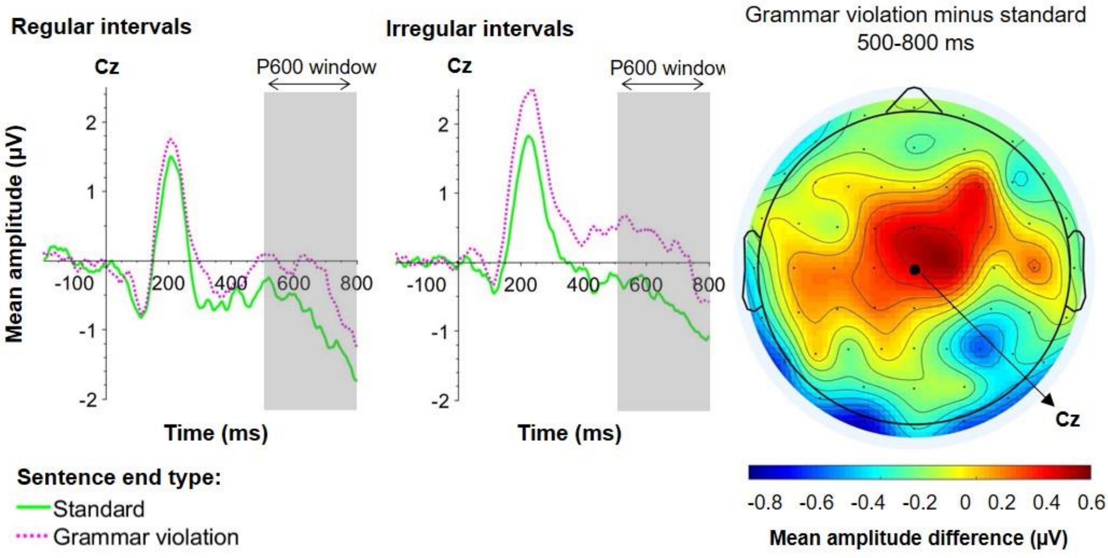


Figure 5.5 – P600 grammar expectancy violation graphs (–200 to 0 ms baseline). Left and middle: Mean amplitude (µV) at Cz between –200 to +800 ms for the two sentence end types: standard word and grammar expectancy violation for the regular 1000 ms (left) and irregular 500/1500 ms (middle) interstimulus intervals. Right: Scalp topography map of the mean amplitude difference (µV) between 500–800 ms after the final (target) word's onset.

5.2.5. Experiment 6: Expertise effects

Analysis of expertise effects, measured by English language experience was not deemed appropriate in Experiment 6, because the numbers of native (*N* = 14) and non-native (*N* = 6) speakers was too unequal. As planned, additional analyses were carried out to investigate whether musical training affected the within- and cross-domain meter expectation ERP effects. A between-subject factor with two levels (amateur musicians, *N* = 12, and non-musicians, *N* = 8) was added to the ANOVA designs for the significant ERP effects. The cut-off was > 1000 hours of practice time for participants to be categorised as amateur musicians (see Table 5.6 for descriptive statistics of each musical training group, and Appendix F for the frequency of accumulated practice times).

Table 5.6 – Descriptive statistics of musical training experience for amateur musicians and non-musicians, Experiment 6

Measure	Amateur musicians (*N* = 12)		Non-musicians (*N* = 8)	
	M (*SD*)	Range	*M* (*SD*)	Range
Number of years practising	9.3 (3.0)	5–14	3.3 (4.6)	0–11
Practice time (hours)	3573 (2211)	1196–9568	201 (229)	0–620
Overall ability rating (1–5)	3.3 (1.0)	1–4	1.8 (0.9)	1–3
Listening per week (hours)	11.8 (6.8)	4–28	12.4 (15.9)	1–50

To investigate whether musical training affected behaviour responses, within-subjects ANOVAs (3 x 2 x 2) were carried out to investigate whether sentence end type (3 levels: standard word, unrelated semantic violation and grammar expectancy violation), interstimulus interval (2 levels: regular 1000 ms or irregular 500/1500 ms) or musical training (2 groups: amateur musicians or non-musicians) affected the percentage of errors participants made on the task or reaction time. There were no main effects or interactions between musical training and sentence end type or interstimulus interval for the percentage of errors or reaction times (all *p*s > .05).

For the ERP analysis, within-subjects ANOVAs were conducted to investigate whether musical training affected the significant ERP effects: the N400 (semantic), P600 (grammar), P2 (meter). There were no main effects or interactions associated with musical training (all *p*s > .05).

5.2.6. Experiment 6: Discussion

The main aim of Experiment 6 was to investigate whether ERP effects for semantic and grammar expectancy violations are affected when presented simultaneously with meter violations, and vice versa. The behavioural results showed no interactions, which was expected because the task was implicit. It was necessary to look at the ERP effects to investigate whether there were interactions for simultaneously presented expectancy violations. The results showed that most language-based ERP effects found in previous chapters were robust. The N400 was found for semantic expectancy violations, the P600 for grammar expectancy violations and the P2 for meter violations. All three effects were not affected by the simultaneous presentation of meter and semantic/grammar expectancy violations. These results suggest that, at least to an extent, neurocognitive processes involved in meter processing are separable from those involved in semantic and grammar processing. The results are discussed in more detail below.

Pre-stimulus effect

As in Experiment 5, the cluster-based permutation statistics identified pre-stimulus effects associated with interstimulus interval. Again, the temporal characteristics of these large clusters overlap with the a-priori time window for the P2 effect, meaning that there are no clear clusters associated with the P2 effect. This pre-stimulus effect was not present in the analysis for the semantic expectancy violations and the grammar expectancy violations. Therefore, although it is still not clear how to interpret the pre-stimulus effect, it appears to be specific to the interstimulus interval manipulations, and does not, for example, seem to be due to a preprocessing issue. All ERP analyses were run with a post-stimulus (0 to +50 ms) baseline in addition to the pre-stimulus (–200 to 0 ms baseline), and all were significant with both baselines. Thus, the analyses support the validity of these ERP effects despite the unexplained pre-stimulus effect.

P2

There was a P2 effect for meter violations in the language domain. The P2 effect is an increase in positive amplitude for irregular interstimulus intervals between the final two words of a sentence compared to regular interstimulus intervals during the a-priori time window (140–260 ms after the final (target) word's onset). This P2 effect is similar to the one identified for meter violations in the language domain in Experiment 5, although in the current experiment, it has a slightly more anterior scalp topography distribution. Previous research has not investigated P2 effects for meter violations in the language domain. While Experiment 5 presented the meter violations in the language domain in a within-subjects design alongside those in the music domain Experiment 6 presented them in the language domain alone. It was deemed important to test whether the P2 effect occurred for language stimuli in a between-subjects design in Experiment 6 in case the within-subject design affected the neurocognitive substrates that were recruited. Together, the current experiments (Experiments 5 and 6) provide a novel contribution to the literature, suggesting that the P2 effect occurs for meter violations in the language domain.

N400

The N400 effect was elicited by semantic expectancy violations, confirming that the N400 effect found in Experiments 2 and 3 is robust. The N400 had similar, centroparietal, scalp topography in all three experiments. This result is consistent with previous studies that report N400 effects for semantic expectancy violations (Gutierrez et al., 2012; Kamp et al., 2015; Kutas & Federmeier, 2010; Kutas & Hillyard, 1983; Moreno & Vázquez, 2011; Nigam et al., 1992; Rommers et al., 2013; Tiedt et al., 2020). As in Experiments 2 and 3, the N400 effect was characterised by an increase in negative amplitude for semantically unrelated words compared to standard words during the a-priori time window (300–450 ms after the final (target) word's onset). The cluster-based permutation statistics identified a significant cluster that corresponded to the N400 effect. As in Experiments 2 and 3, this cluster confirmed that the a-priori time window of 300–450 ms is suitable for investigating the N400 effect, as the temporal extent of the cluster fell mainly within this time window (289–445 ms).

P600

A P600 effect was elicited by grammar expectancy violations. It has similar characteristics, such as central scalp topography, in Experiments 2, 3 and 6. The P600 has also been reported for grammar expectancy violations in several previous studies (Hahne et al., 2012; Mehravari et al., 2015; Osterhout & Holcomb, 1992). The P600 effect was characterised by a central increase in positive amplitude for grammar expectancy violations compared to standard words between 500–800 ms after the final (target) word's onset (the a-priori time window). Unlike Experiments 2 and 3, the cluster-based permutation statistics did not identify any clusters associated with grammar expectancy violations. This could be due to the more complex experimental design (as grammar expectancy violations were combined with meter violations, instead of being presented on their own). The significant clusters associated with interstimulus intervals were large, therefore smaller clusters (like that associated with the P600) were unlikely to be detected.

Interactions between interstimulus interval and sentence end type

There were no interactions between interstimulus interval and sentence end type, for any of the ERP effects (P2 for meter violations, N400 for semantic violations or P600 for grammar violations). Therefore, the results suggest that meter processing is, at least to an extent, separable from semantic and grammar processing. Some previous studies have reported reduced N400 and P600s when semantic and grammar expectancy violations were presented simultaneously with irregular meter (Canette et al., 2020; Rothermich et al., 2010), others report increased effects (Li et al., 2019; Schmidt-Kassow & Kotz, 2009) and others, like the current experiment, reported no difference (Magne et al., 2007). Interactions in the N400 semantic and P600 grammar violation effects require further investigation, perhaps with several types of meter and rhythm violation presented within the same study, so that more confident conclusions about neurocognitive overlap between meter processing and semantic and grammar processing can be made. Although further testing is required, the current evidence suggests that the mechanisms involved in processing irregular inter-stimulus intervals that are indicated by the P2 effect are specialised for meter processing.

Expertise effects

As in the previous experiments that adopted implicit tasks in this thesis (Experiments 3 and 5), there were no expertise effects associated with any type of expectancy violations effects (behavioural or ERP) in Experiment 6. Therefore, there was no evidence of cross-domain expertise effects from musical expertise to processing of language-based expectancy violations, which may have indicated shared neurocognitive substrates between language and music domains. Like in the previous experiments, it is possible that the distinction between amateur musicians and non-musicians (> 1000 hours for amateur musicians, < 1000 hours for non-musicians) did not create a large enough difference in musical training between groups that would be needed to detect expertise effects. This idea is discussed in more detail in section 6.1.6.

Conclusion

Experiment 6 showed that the P2 effect for meter violations in the language domain, the N400 effect for semantic expectancy violations and the P600 effect for grammar expectancy violations are robust. There were no interactions in these ERP effects when meter and semantic, or grammar expectancy violations were presented simultaneously. Overall, the results support the hypothesis that the proposed “domain-specialised” (N400 for semantic expectancy violations and P600 for grammar expectancy violations) and “domain-general” (P2 for meter violations) neurocognitive substrates exist and are, at least to an extent, separable. The next experiment,

Experiment 7, investigates whether there are interactions in ERP effects for meter and harmony expectancy violations when they are presented simultaneously.

5.3. Experiment 7: Music domain experiment – simultaneous presentation of meter and harmonic expectancy violations

5.3.1. Experiment 7: Rationale and aims

As discussed in detail in Chapter 1, section 1.3.4, the current evidence suggests that there might be separable neurocognitive substrates for meter and some aspects of music processing. Most previous evidence suggests a separation between neurocognitive substrates involved in meter and pitch processing (Midorikawa et al., 2003; Peretz & Coltheart, 2003; Peretz & Hyde, 2003; Peretz & Zatorre, 2005; Phillips-Silver et al., 2013; Vignolo, 2003). However, there is also evidence focussing on harmony. In contrast to pitch, previous evidence suggests that there could be shared neurocognitive substrates for meter and harmony processing. For example, harmonically expected target chords have been judged as being better fitting (Schmuckler & Boltz, 1994) and more complete (Tillmann & Lebrun-Guillaud, 2006) when they were played with a regular meter compared to an irregular meter. Likewise, Jung et al. (2015) found that a metrically regular presentation of harmonic expectancy violations decreased reaction times to their correctness judgements, compared to an irregular meter.

Zhang et al. (2019) presented chord progressions in which the final two chords were harmonically expected or created harmonic suspense. The chord progressions either provided a sense of harmonic closure (the final two chords were V–I, a perfect cadence), or harmonic suspense (the final two chords were I–IV). They were presented either in a regular meter in relation to the rest of the chord progression (one chord every 600 ms) or slightly earlier (524 ms after the penultimate chord). The N500 effect was only found for harmonic expectancy violations when the meter was regular, and not when the meter was irregular. Therefore, there is previous evidence of an interaction in an ERP effect associated with harmonic violations when they were simultaneously presented with meter violations. In Zhang et al.'s (2019) study, is not clear what the effects were for meter violations separately. Therefore, it is possible that there were two separate components with overlapping scalp topography: one for harmonic tension (the N500) and one for meter violations. If this were the case, the N500 effect might not be detected in the simultaneous presentation design, but the results would not necessarily suggest shared neurocognitive substrates between meter and harmonic processes. With this in mind, in Experiment 7, ERP effects are presented separately for meter and harmonic violations, as well as testing for interactions.

Experiment 7 was designed to investigate whether ERP effects associated with meter violations (N1 and P2) and harmonic expectancy violations (P300) interact when the two violation types are presented simultaneously. While Zhang et al. (2019) investigated interactions between meter and harmonic processes for chords that created harmonic tension, eliciting the N500 effect, Experiment 7 aimed to investigate interactions between meter and harmonic processes for harmonic violations, which elicit the P300 effect. Based on the results of Zhang et al. (2019), it was predicted that irregular interstimulus intervals would reduce the P300 harmonic expectancy violation effect.

Janata et al. (2002) found that temporal asynchrony judgements for a target chord were less accurate when the chord was preceded by harmonically unrelated chords compared to when the chord was preceded by chords in the same key (harmonically expected).Based on these results, it was tentatively predicted that the N1 and P2 meter violation effects would be reduced when meter violations were simutaneously presented with harmony violations.

5.3.2. Experiment 7: Method

Ethics

The University Teaching and Research Ethics Committee at The University of St Andrews approved this study, code PS13256 (see Appendix P).

Participants

Twenty-three university students volunteered to participate in this study. The data of three participants were not used for analysis due to technical issues during EEG recording. The final *N* = 20 (4 male, Age: *M* = 22 years, *SD* = 5 years, range = 18–38 years). Descriptive statistics of participants' English language and musical training experience are in Table 5.7. Each testing session took 120 minutes, and participants were reimbursed £10. People could not sign up to participate in the study if they had participated in Experiment 1, 2, 3, 4 or 5.

Table 5.7 – Descriptive statistics of participants' English language and musical training experience, Experiment 7.

Experience	Measure	*M* (*SD*)	Range
English language	Native English speaking	Native: 9, Non-native: 11	
	Onset (age in years)	2.7 (3.0)	0–8
	English ability (rating 1–10)	9.4 (0.9)	7–10
	% English spoken in the past year	67.5 (27.2)	5–100
Musical training	Number of years practising	8.6 (5.4)	0–19
	Practice time (hours)	1630 (1504)	0–4498
	Overall ability rating (1–5)	3.1 (1.1)	1–4
	Listening per week (hours)	18.5 (13.8)	1–50

Apparatus

For details of the computer set up and response keys, see section 2.2.2 Experiment 1: Method (Apparatus). For details of the EEG system, see section 3.2.1 Experiment 2: Method (Apparatus).

Materials

Chord progressions

There were 192 chord progression presentations in the main experiment. The 24 chord progression stems from Experiments 2, 3, 4 and 5 (adapted from chord progressions presented in Bach-Gesellschaft (1892)) were used. For this experiment, they were presented with two chord progression end types (tonic (I) and unrelated harmony) and two types of the interstimulus interval between the final two chords (regular 1000 ms and irregular 500/1500 ms). For the irregular interstimulus intervals, half were short (500 ms) and half were long (1500 ms). Therefore, each of the 24 chord progression stems was presented four times (see Table 5.8). These 96 chord progressions were presented twice: once played by the piano and once played by the organ. For each chord progression (played by both piano and organ), participants were asked whether it was played by the “Piano?” or “Organ?”, to which they answered “Yes” or “No”.

Table 5.8 – The four presentation types of the 24 chord progression stems.

Chord progression end type		Interstimulus interval
Tonic (I)	and	Regular (1000 ms)
Tonic (I)	and	Irregular (500/1500 ms)
Unrelated harmony	and	Regular (1000 ms)
Unrelated harmony	and	Irregular (500/1500 ms)

Additional materials

See section 3.2.1 Experiment 2: Method (Materials) for details of the Language Experience Questionnaire. See section 2.2.2 Experiment 1: Method (Materials) for details of the Musical Training Questionnaire.

Procedure

After participants read the information sheet and signed the consent form, they were prepared for EEG recording. Participants sat in an electrically shielded booth. Auditory material was presented over loudspeakers, and the questions (“Piano?” and “Organ?”) were presented visually on a computer screen.

There were 12 practice trials, with three presentations of each of the four presentations (refer to Table 5.8). 1000 ms after the final word or chord onset, participants were asked whether the progression was played by the “Piano?” or “Organ?”. They responded “Yes” or “No“ using the response keys. Participants chose whether they answered “Yes” with the right hand and “No” with the left (*N* = 13), or vice versa (*N* = 7).

After the practice trials, the experimental trials were presented. There were eight blocks of 24 experimental trials. The sentences were presented in random order across the eight blocks. Participants were able to take breaks after each block. After the task was completed, participants were asked to fill in the Language Experience Questionnaire and the Musical Training Questionnaire.

Analysis

For details of the analysis and graphing software and statistical correction procedures used, see section 2.2.2 Experiment 1: Method (Analysis). For details of EEG analysis procedures, see section 3.2.1 Experiment 2: Method (Analysis).

Trial inclusion

As in the previous experiments with the implicit tasks, only trials with correct responses given between 100–5000 ms after the onset of the question prompt were considered for analysis. Trials were also excluded with artefact rejection using the moving peak-to-peak window method after ICA correction in ERPLAB. Table 5.9 shows the average number of trials that remained per condition after the inclusion criteria were applied. All end types had an average of 42 or more trials included in the final dataset, so based on Luck’s (2014) recommendations, the final dataset had an acceptable signal-to-noise ratio.

Table 5.9 – Mean (*SD*) number of trials included in the analysis after the inclusion criteria were applied 7. The maximum possible number of trials for each of the four presentation types = 48.

		Mean number of trials left (*SD*)	
Chord progression end type	Interstimulus interval	Behaviour	Artefact rejection
Tonic (I)	Regular (1000 ms)	46.1 (2.2)	43.3 (4.3)
Tonic (I)	Irregular (500/1500 ms)	46.1 (1.7)	43.2 (4.1)
Unrelated harmony	Regular (1000 ms)	46.3 (2.0)	43.2 (3.8)
Unrelated harmony	Irregular (500/1500 ms)	45.4 (2.8)	42.3 (4.8)

Analysis of ERP effects

Cluster-based permutation statistics were run between 0–800 ms relative to the onset of the final chord, with the Factorial Mass Univariate ERP Toolbox (Fields, 2017; Groppe et al., 2011). The ANOVA analyses parameters used for the ERP effects are the same as in previous chapters to allow direct comparison (see Table 5.10). Refer to sections 3.2.1 Experiment 2: Method 4.3.2 (Analysis) and 3.3.2, Experiment 3: Method (Analysis) for justification of the P300 analysis choices and section 4.3.2 Experiment 5: Method (Analysis) for justification of the N1 and P2 analysis choices.

Table 5.10 – A-priori selected electrodes and time windows for analysis the N1, P2 and P300 components

ERP component	Electrode	Time (ms)	Violation type
N1	Fz	80–140	Meter
P2	Pz	140–160	Meter
P300	Fz	250–400	Harmonic

To control the pre-stimulus effect that was discovered in Experiment 5, a post-stimulus baseline (0 to +50 ms) analysis was included in addition to the usual pre-stimulus (–200 to 0 ms) baseline analysis, in attempt to test the validity of effects associated with the onset of the final chord (the target chord).

Analysis of expertise effects

As in the previous experiments, to retain statistical power to detect ERP effects, the analyses first included all participants. Then, for any significant ERP effects, two between-subjects factors, English language experience (2 levels: native and non-native speakers) and musical training (2 levels: amateur musicians and non-musicians) were added to the ANOVAs to test for expertise effects.

5.3.3. Experiment 7: Behaviour results

Percentage of errors

Two 2 x 2 within-subjects ANOVAs investigated whether chord progression end type (2 levels: tonic (I) and unrelated harmony) or interstimulus interval (2 levels: regular (1000 ms) and irregular (500/1500 ms) affected the percentage of errors or reaction times. There were no main effects or interactions of chord progression end type or interstimulus interval on the percentage of errors participants made (*ps* > .05), see Figure 5.6.

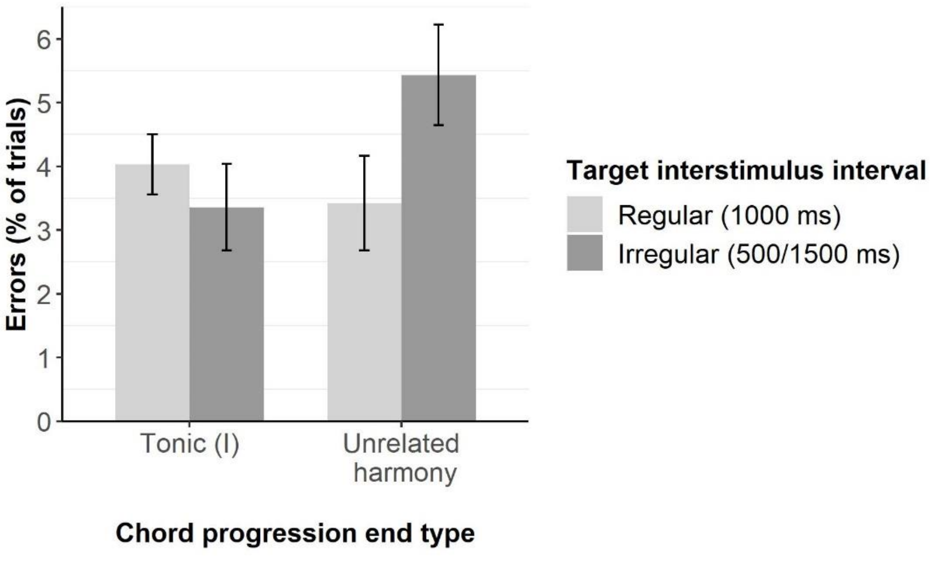


Figure 5.6 – Mean % of errors for chord progression end types and interstimulus intervals. Within-subject SE bars are included.

Reaction time

Chord progression end type affected reaction time *F*(1, 19) = 10.48, *p* = .004, ηp^2 = 0.36. Reaction time was faster for the tonic (I) chord *(M* = 794 ms, *SD* = 74 ms) compared to the unrelated harmony chord *(M* = 841 ms, *SD* = 75 ms), see Figure 5.7, left panel. There were no other significant effects or interactions for reaction time (all *p*s > .05).

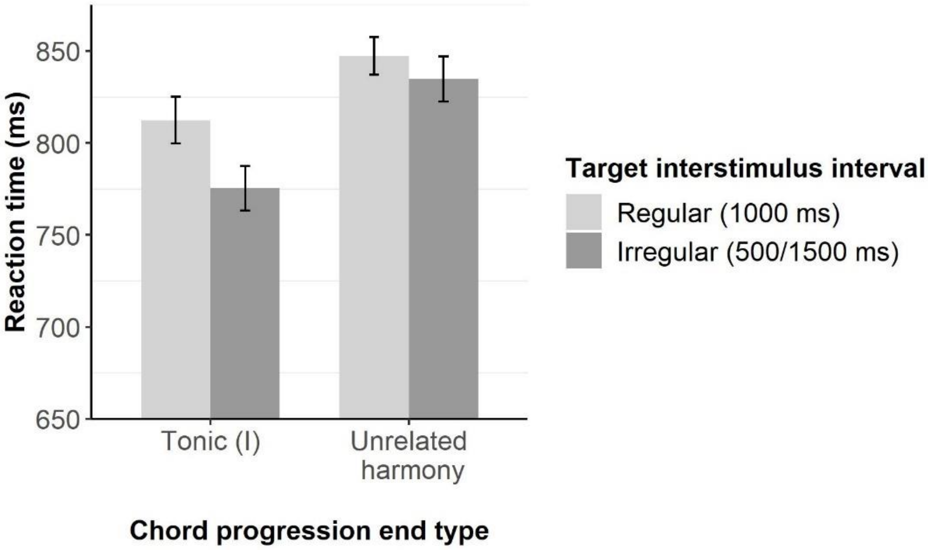


Figure 5.7 – Mean reaction time (ms) for chord progression end types and interstimulus intervals. Within-subject SE bars are included.

5.3.4. Experiment 7: ERP results

As in Experiments 5 and 6, there was an additional pre-stimulus effect for the effect of interstimulus interval in Experiment 7, which was captured by two clusters in the within-subjects cluster-based permutation statistics (see Table 5.11), carried out between 0–800 ms after the final (target) chord's onset. This effect affected the pre-stimulus baseline (–200 to 0 ms) for the interstimulus interval manipulation. As planned, all significant ERP effects for the interstimulus interval, analyses were also run with a post-stimulus baseline (0 to +50 ms) to test their validity. Both baseline analyses are referred to in the results section. The results of the 0 to +50 ms baseline *F* tests are in Appendix M. There were no significant clusters associated with harmonic expectancy violations (tonic (I) or unrelated harmony) and, crucially, no significant interactions between harmonic expectancy violations and meter violations. Therefore, these clusters appear to be specific to the inter-stimulus interval manipulation.

Table 5.11 – Cluster-based permutation statistics results for the effect of interstimulus interval (data collapsed across harmonic end types).

Mass	*p*	Electrodes	Temporal peak (ms)	Spatial peak	Extent (ms)
4652	<.001	TP7, P1, P5, P7, PO7, O1, Oz, Pz, F4, FC4, FC2, C4, C6, T8, TP8, CP6, CP4, CP2, P2, P4, P6, P8, PO8, PO4, O2	156	CP4	Start: 16 End: –
3212	.001	Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C3, T7, TP7, CP3, P7, Fpz, Fp2, AF4, Fz, F2, F4, F6, FC4, FC2	141	F3	Start: 23 End: –

To investigate ERP effects, within-subjects 2 x 2 within-subjects ANOVAs were conducted for each ERP effect previously found for harmonic expectancy violations in Experiments 2 and 3 (P300) and meter violations in Experiment 5 (N1 and P2). These analyses were designed to investigate whether chord progression end type (2 levels: tonic (I) and unrelated harmony) or interstimulus interval (2 levels: regular 1000 ms and irregular 500/1500 ms) elicited these effects and, crucially, to investigate potential interactions.

N1

Interstimulus interval manipulations affected the mean amplitude at Fz in the N1 time window (80–140 ms after the final (target) chord's onset) *F*(1, 19) = 5.03, *p* = .037, ηp^2 = 0.21, see Figure 5.8. The mean amplitude was more negative for the irregular (500/1500ms) interstimulus interval (*M* = –0.51 µV, *SD* = 0.17 µV) compared to the regular interstimulus interval (*M* = –0.01 µV, *SD* = 0.18 µV). This effect was also significant with the 0 to +50 ms baseline, *F*(1, 19) = 7.02, *p* = .016, ηp^2 = 0.27, see Appendix M. There was no effect of chord progression end type, *F*(1, 19) = 0.52, *p* = .478, ηp^2 = 0.03, and no interactions between interstimulus interval and chord progression end type, *F*(1, 19) = 0.62 *p* = .443, ηp^2 = 0.03, see Appendix N for the 0 to +50 ms baseline interaction result (*p* > .05).

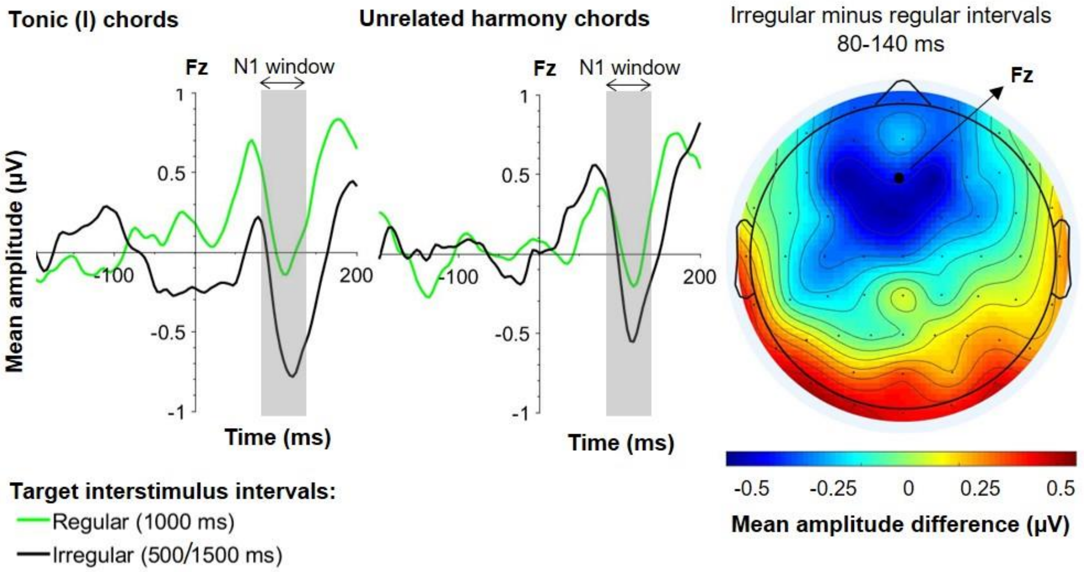


Figure 5.8 – N1 language meter violation graphs (–200 to 0 ms baseline). Left and middle: ERP graphs at Fz between –200 to +200 ms for the regular and irregular interstimulus intervals, separate for the two chord progression end types: tonic (I) (left) and unrelated harmony (middle). Right: Scalp topography map of the mean amplitude difference (µV) between 80–140 ms after the final (target) chord's onset.

P2

Interstimulus interval manipulations affected the mean amplitude at Pz in the P2 time window (140–260 ms after the final (target) chord's onset) *F*(1, 19) = 10.34, *p* = .005, ηp^2 = 0.35, see Figure 5.9. The mean amplitude was more positive for the irregular (500/1500 ms) interstimulus interval (*M* = 0.74 µV, *SD* = 0.20 µV) compared to the regular interstimulus interval (*M* = 0.27 µV, *SD* = 0.17 µV). With the 0 to +50 ms baseline, this effect was not significant, but the trend remained, *p* = .086, see Appendix M. There was no effect of chord progression end type, *F*(1, 19) = 0.82, *p* = .376, ηp^2 = 0.04, and no interactions between interstimulus interval and chord progression end type, *F*(1, 19) = 0.31, *p* = .587, ηp^2 = 0.02, see Appendix N for the 0 to +50 ms baseline interaction result (*p* > .05).

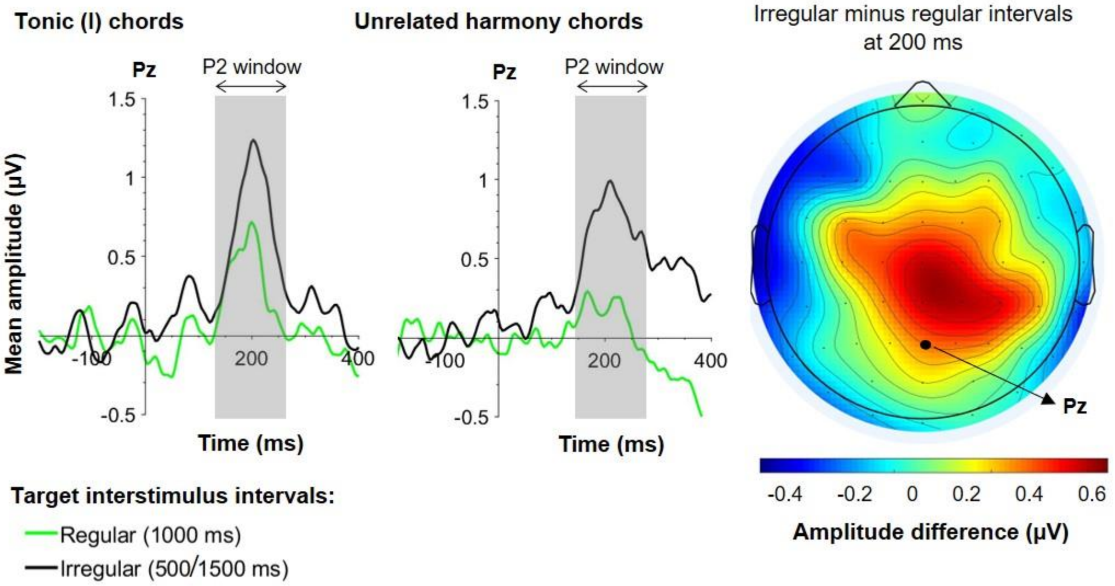


Figure 5.9 – P2 language meter violation graphs (–200 to 0 ms baseline). Left and middle: ERP graphs at Pz between –200 to +400 ms for the regular and irregular interstimulus intervals, separate for the two chord progression end types: tonic (I) (left) and unrelated harmony (middle). Right: Scalp topography map of the mean amplitude difference (µV) at 200 ms after the final (target) chord's onset.

P300

A trend (not significant at the $p < .05$ level) suggested that the different harmonic endings had some effect on the mean amplitude at Fz in the P300 time window (250–400 ms after the final (target) chord's onset) $F(1, 19) = 4.10$, $p = .057$, $\eta p^2 = 0.35$, see Figure 5.10. The mean amplitude tended to be more positive for the unrelated harmony chords ($M = –0.05$ µV, $SD = 0.24$ µV) compared to the tonic (I) chords ($M = –0.52$ µV, $SD = 0.20$ µV). This effect was also significant with the 0 to +50 ms baseline, see Appendix M. Within-subjects cluster-based permutation statistics did not identify any significant clusters associated with harmonic expectancy violations. Crucially, there were no interactions between chord progression type and interstimulus interval, $F(1, 19) = 0.96$, $p = .340$, $\eta p^2 = 0.05$, see Appendix N for the 0 to +50 ms baseline interaction result ($p > .05$).

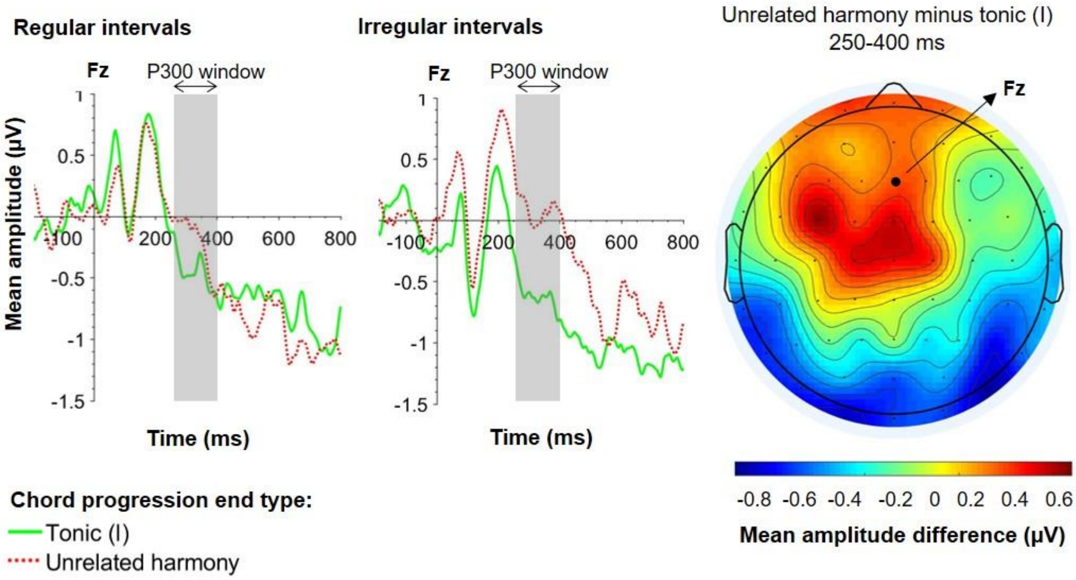


Figure 5.10 – P300 harmonic expectancy violation graphs (–200 to 0 ms baseline). Left and middle: Mean amplitude (µV) at Fz for the tonic(I) and unrelated harmony chord progression end types for the regular 1000 ms (left) and irregular 500/1500 ms interstimulus intervals. Right: Scalp topography map of the mean amplitude difference (µV) between 250–400 ms after the final (target) chord's onset.

5.3.5. Experiment 7: Expertise effects

As planned, additional analyses were carried out to investigate whether musical training or English language experience affected within- or cross-domain effects. Two between-subject measures were included in the ANOVA designs: English language experience (2 groups: 9 native and 11 non-native English speakers) and musical training (2 groups: twelve amateur musicians and eight non-musicians). For descriptive statistics of English language experience for native and non-native speakers, see Table 5.12. For musical training experience, the cut-off was > 1000 hours of practice time for participants to be categorised as amateur musicians and < 1000 hours of practice time for non-musicians. See Table 5.13 for descriptive statistics of each musical training group, and Appendix F for the frequency of accumulated practice times.

Table 5.12 – Descriptive statistics of English language experience for native and non-native speakers, Experiment 7

Measure	Native speakers (*N* = 9)		Non-native speakers (*N* = 11)	
	M (*SD*)	Range	*M* (*SD*)	Range
Onset learning English (years)	0.0 (0.0)	0	5.0 (2.2)	2–8
Overall ability rating (1–10)	10.0 (0.0)	10	8.9 (0.9)	7–10
% time speaking English	86 (17)	60–100	52 (25)	5–80

Table 5.13 – Descriptive statistics of musical training experience for amateur musicians and non-musicians, Experiment 7

Measure	Amateur musicians (*N* = 12)		Non-musicians (*N* = 8)	
	M (*SD*)	Range	*M* (*SD*)	Range
Number of years practising	12.1 (2.2)	8–19	3.3 (3.5)	0–10
Practice time (hours)	2552 (1248)	1040–4498	247 (230)	0–624
Overall ability rating (1–5)	3.6 (0.7)	2–4	2.4 (1.3)	1–4
Listening per week (hours)	16.6 (9.1)	2–28	32.4 (19.2)	1–50

To investigate whether musical training affected behaviour responses, mixed measures ANOVAs (2 x 2 x 2 x 2) were carried out to investigate whether chord progression end type (3 levels: tonic (I) and unrelated harmony), interstimulus interval (2 levels: regular 1000 ms or irregular 500/1500 ms), musical training (2 groups: amateur musicians or non-musicians) or English language experience (2 groups: native and non-native) affected the percentage of errors participants made on the task or reaction time. There were no interactions between musical training and sentence end type or interstimulus interval for the percentage of errors or reaction times (all *p*s > .05).

For the ERP analysis, within-subjects ANOVAs were conducted to investigate whether musical training or English language experience affected the significant ERP effects: the P300 (harmony), N1 (meter) or P2 (meter). There were no main effects or interactions associated with musical training or English language experience (all *ps* > .05).

5.3.6. Experiment 7: Discussion

The main aim of Experiment 7 was to investigate whether ERP effects for harmonic expectancy violations are affected when they are presented simultaneously with meter violations and vice versa. The behavioural results showed no interactions, which was expected because the task was implicit. It was necessary to look at the ERP effects to investigate whether there were interactions for simultaneously presented expectancy violations. The results show that the N1 and P2 effects occurred for meter violations, and the P300 effect occurred for harmonic expectancy violations. These results demonstrate the reliability of these ERP effects, which were also found in Experiments 2, 3 and 5. Crucially, these ERP effects were not affected by the simultaneous presentation of meter and harmonic expectancy violations. Overall, these results support the idea that neurocognitive substrates involved in meter processing are at least somewhat separable from those involved in harmony processing.

Pre-stimulus effect

As in Experiment 5, the cluster-based permutation statistics identified pre-stimulus effects associated with interstimulus interval. Again, the temporal characteristics of these large clusters overlap with the a-priori time window for the N1 and P2 effects, and there are no clear clusters associated with these effects. This pre-stimulus effect was not present in the analysis for the harmonic expectancy violations. Therefore, although it is still not clear how to interpret the pre-stimulus effect, it appears to be specific to the interstimulus interval manipulations. All ERP analyses were run with a post-stimulus 0 to +50 ms baseline in addition to the pre-stimulus –200 to 0 ms baseline. All ERP effects were significant with both baselines. As a result, there is evidence to support that the reported ERP effects associated with metric violations (the N1 and P2) are valid, despite this unexplained pre-stimulus effect.

N1

As in Experiment 5, an N1 effect was elicited by meter violations in the music domain. There was an increase in negativity for irregular interstimulus intervals between the final two chords compared to regular interstimulus intervals. The N1 effect was frontally distributed and occurred between 80–140 ms after the final (target) chord's onset (the a-priori time window). Similar N1 effects have been reported for different types of meter violations in the music domain in previous studies (Fitzroy & Sanders, 2015; Habibi et al., 2014; Kober et al., 2015). The results of Experiment 5 and the current experiment, Experiment 7, supports that the N1 effect is a reliable effect for meter violations in the music domain.

P2

As in Experiment 5, there was a P2 effect for meter violations in the music domain. It was characterised by an increase in positivity for irregular interstimulus intervals between the final two chords of the chord progressions compared to regular interstimulus intervals in the a-priori time window (140 –260 ms after the final (target) chord's onset). Similar P2 effects are reported in previous studies, which have created meter violations in the music domain in various ways (Neuhaus & Knösche, 2008; Raij et al., 1997). Therefore, the P2 effect identified in the current experiments (Experiments 5 and 7) appears to be a reliable effect for meter violations in the music domain.

P300

The P300 effect for harmonic expectancy violations was not statistically significant but a trend (*p* = .057), alongside the ERP and scalp topography maps, suggest harmonic expectancy violations elicited a P300 effect. It was characterised by an increase in positivity for unrelated harmony chords compared to tonic (I) chords in the a-priori time window (250–400 ms). This P300 effect is similar to those found in Experiments 2 and 3, suggesting that it is a reliable harmonic expectancy violation effect for the harmonic expectancy violations that were chosen due to the results of Experiment 1. In all current three experiments, it had similar, frontal, scalp topography. Previous studies have reported similar P300 effects for harmonic violations (Beisteiner et al., 1999; Carrión & Bly, 2008; Janata, 1995; Loehr et al., 2013).

Unlike in Experiments 2 and 3, the P300 was not statistically significant at the *p* < .05 level in the current experiment, and the within-subjects cluster-based permutation statistics did not identify any clusters associated with harmonic expectancy violations. This could be because the experimental design was more complicated (as harmonic expectancy violations were combined with meter violations, instead of being presented on their own), and this may have reduced their saliency. The clusters associated with interstimulus intervals are likely to be due to an unexplained pre-stimulus effect, and, therefore, smaller clusters, like the one associated with the P300, were unlikely to be detected.

Interactions between interstimulus interval and harmonic end type

There were no interactions between interstimulus interval and harmonic end type for any of the ERP effects (N1 and P2 for meter violations, or P300 for harmonic expectancy violations). Consequently, the neurocognitive substrates involved in meter and harmony processing appear to be, at least to an extent, separable. This finding could be considered alongside the model of the modularity of music processing proposed by Peretz and Coltheart (2003), which suggests that pitch and meter information are processed separately. It is possible that this model could extend to from pitch to some types of harmony processing. Similar to how Besson et al. (1998) proposed an additive model of semantic and harmony processing when there were no interactions for simultaneous presentations of semantic and harmonic expectancy violations in the N400 and P300 effects, based on the current results, one might propose an additive model for meter and harmonic processing, at least in terms of the neurocognitive substrates indicated by the current ERP effects.

While Zhang et al. (2019) report that the N500 effect elicited by harmonically suspenseful stimuli was affected by simultaneous presentation with meter violations, the current study suggests that the P300 effect elicited by harmonically unexpected chord progression endings that do not create suspense is not affected in the same way. Perhaps harmonically suspenseful stimuli interact with meter violations due to the temporal nature of both types of stimuli. For harmonic tensions, participants may anticipate that a resolution will occur over time. Therefore, the neurocognitive substrates indicated by the N500 effect may be involved in temporal processes, which could explain the interaction reported by Zhang et al. (2019). In contrast, for the current unexpected harmonic endings, participants do not anticipate further stimuli. This hypothesis could explain the differences between Zhang et al.'s (2019) results and the results of the current study and it requires further investigation.

Expertise effects

There were no associations between musical training or English language experience associated with the N1, P2 or P300 effects. Therefore, there was no evidence of within- or cross-domain expertise effects. It is possible that the distinctions between native and non-native English speakers and amateur musicians and non-musicians did not distinguish proficiencies enough to detect expertise effects. This idea is discussed in more detail in section 6.1.6.

Conclusion

Overall, Experiment 7 shows that the N1 and P2 effects for meter violations in the music domain and the P300 effect elicited by harmonic expectancy violations are robust. There were no interactions in these ERP effects when the meter and harmonic expectancy violations were presented simultaneously. Therefore, the results support the hypothesis that the neurocognitive substrates indicated by the P300 for harmonic expectancy violations and those indicated by the N1 and P2 for meter violations are, at least to an extent, separable.

5.3.7. Experiments 6 and 7: Test of domain-general ERP effects

After analysing the results of Experiments 6 and 7 separately, a cross-experiment analysis was conducted to further investigate the case for domain-general ERP effects associated with meter processing. In both Experiments 6 and 7 a similar P2 effect was found in both language and music domains, see Figure 5.11. However, if there were any interactions between experiments (Experiment 6: Language and Experiment 7: Music) and interstimulus interval, the case for a domain-general P2 effect would be weakened. Therefore, a 2 x 2 mixed ANOVA was carried out to investigate whether the P2 effect was affected by the interstimulus interval (2 levels, within-subjects: regular 1000 ms and irregular 500/1500 ms) or task (2 levels, between-subjects: language task and music task). For both –200 to 0 ms and 0 to +50 ms baselines, the effect of the interstimulus interval was significant, showing the P2 meter violation effect (see Appendix O). There was also a main effect of task $F(1, 38) = 4.63$, $p = .038$, $\eta p^2 = 0.11$. During the P2 time window, the mean amplitude was generally more positive in the language task ($M = 1.02$ µV, $SD = 0.17$ µV) compared to the music task ($M = 0.51$ µV, $SD = 0.17$ µV). This effect is of little interest, as it does not relate to the meter violations, and, therefore, it will not be discussed further.

Crucially, there was no interaction between interstimulus interval and task $F(1, 38) = 2.40$, $p = .130$, $\eta p^2 = 0.06$. There was also no interaction in the 0 to +50 ms baseline analysis (see Appendix O). These findings support the interpretation that the P2 meter violation effect could be thought of as a domain-general effect in both language and music domains. It could therefore indicate domain-general neurocognitive substrates for meter processing.

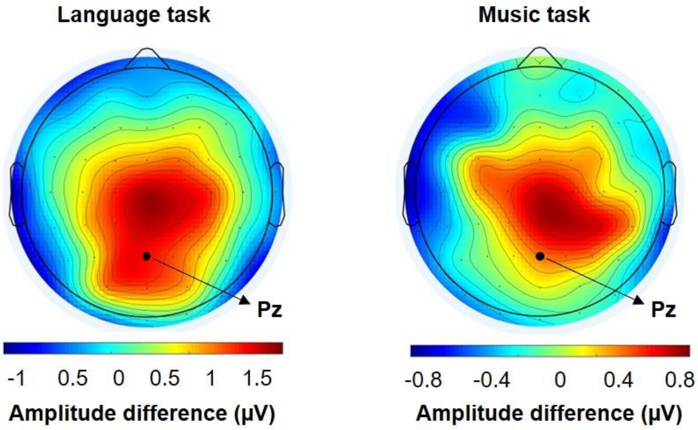


Figure 5.11 – P2 Scalp topography maps from the language task, Experiment 6 (left) and music task, Experiment 7 (right) (-200 to 0 ms baselines). The plots show the amplitude difference for the irregular interstimulus intervals (average of short and long) minus that for the regular interstimulus intervals at 200 ms. These scalp maps are also presented in the results sections above and are presented again here to facilitate comparison. Please note: these scalp maps are on different scales.

5.4. Conclusion

The main aim of the experiments presented in this chapter (Experiments 6 and 7) was to investigate whether the ERP effects identified in previous experiments were robust and, crucially, whether there is evidence of neural separation or interactions between ERP effects for meter violations and ERP effects for semantic, grammar and harmonic processes when they are presented simultaneously.

First, the results showed robust replicability of the ERP effects reported in previous chapters of this thesis, for semantics (N400), grammar (P600) harmony (P300), and meter violations (N1 in the music domain, P2 in the language and music domains).

Second, the results revealed no interactions between meter processes and semantic, grammar or harmonic processes in any ERP effects. Altogether, these results support the existence of specialised neurocognitive substrates for semantics and grammar processing in the language domain, and for harmony processing in the music domain. The results also appear to indicate domain-general mechanisms that are specialised for meter processing in both domains.

6. Chapter 6 – Discussion

This thesis aimed to contribute to the age-old interdisciplinary debate over the extent of neurocognitive overlap between language and music domains. To achieve this, ERP effects for semantic, grammar, harmony and meter expectancy violations were investigated. The main research questions were as follows:

- Is there evidence that ERP effects for semantic, grammar and harmonic expectancy violations indicate separable neurocognitive substrates? (Chapter 3)
- Are there similar ERP effects for meter violations in language and music domains? (Chapter 4)
- Do ERP effects for meter violations interact with those for semantic, grammar or harmonic violations? (Chapter 5)
- Are there cross-domain effects of language or musical expertise in any ERP effects? (Chapters 3, 4 and 5)

In this chapter, the findings of this thesis are summarised and discussed alongside its limitations, directions for future research and potential practical implications.

6.1. Thesis findings

6.1.1. Summary of the main findings in each chapter

The aim of Chapter 2 was to create expectancy violations that would facilitate active comparisons of ERP effects between language and music domains. The anticipated sentence endings were high-cloze words (> 90% cloze probability) adapted from Block and Baldwin's (2010) study. Two levels of semantic expectancy violations were created: related semantic violation (the meaning would be unlikely in the context of the sentence but not impossible) and unrelated semantic violation (the meaning was highly unlikely, or impossible). Grammar violations were the high-cloze probability words but in incongruent word forms (e.g., unexplained mixtures of tenses, singular versus plural words, or verbs versus nouns). Harmonic violations were defined as chords that were not the anticipated tonic (I) chord. These stimuli were based on previous literature that had reported the N400 ERP effect for semantic violations, the P600 for grammar violations and the P300 for harmonic violations. The previous literature on semantic and grammar violations is well-validated, and it was not deemed necessary to test the expectancies of the different sentence end types at this stage of the thesis. On the other hand, the literature currently lacks systematic investigations into the perceived relative expectancies of different chords in place of an anticipated tonic (I) chord. Therefore, Experiment 1 was designed to identify such chords that participants reliably rate "Expected" or "Unexpected". Three levels of expectancy were identified – expected (tonic(I)), ambiguous (related harmony) and unexpected (unrelated harmony). These stimuli were used to test ERP effects associated with harmonic expectancy violations in later experiments.

In Chapter 3, Experiments 2 and 3 aimed to test the distinctness of ERP effects associated with semantic, grammar and harmonic expectancy violations. As predicted, an N400 effect was elicited by semantic violations, a P600 for grammar violations and a P300 for harmonic violations. These effects were robust for the more severe violation types in both Experiment 2 (when the expectancy violations were task-relevant) and in Experiment 3 (when they were task-irrelevant). In Experiment 2, there were unforeseen ERP components in the music task, which appeared to be associated with task difficulty. These components were not present in Experiment 3. Based on these findings, all future experiments used implicit tasks.

Previously, ERP effects for meter violations have not been actively compared between language and music domains. In Chapter 4, Experiment 4 identified interstimulus interval manipulations that were comparably rated as "Expected" or "Unexpected" in language and music domains. Experiment 5 aimed to investigate ERP effects associated with these meter violations. While the evidence for or against an N1 effect in the language domain was not clear, there was clear evidence of an N1 effect in the music domain. Crucially, P2 effects were elicited for meter violations in both language and music domains.

In Chapter 5, semantic, grammar and harmonic expectancy violations were presented simultaneously with meter violations. This design allowed for the investigation of two things—first, to test the reliability of the ERP effects presented in previous chapters and, second, to test for interactions in ERP effects. The results showed that the ERP effects presented in previous chapters were robust and reliable. Crucially, there were no interactions between the ERP effects for meter violations (N1 music domain, P2 language and music domains) and those for semantic (N400), grammar (P600) or harmony (P300) violations.

6.1.2. Domain-specialised ERP effects

The experiments presented in this thesis indicate distinct ERP effects for semantic (N400), grammar (P600) and harmonic violations (P300), supporting the hypothesis that there are domain-specialised neurocognitive substrates involved in these processes. These effects were present for the most severe (“unrelated”) violation types, for both within-subjects designs, when these violation types were presented independently, and between-subjects designs, when these violation types were presented simultaneously with meter violations. The within-subjects designs enabled novel insights into their distinctness as differences in scalp topography could not be attributed to differences in factors such as the analysis protocol. It also enabled direct tests of whether ERP effects elicited by one type of violation were also elicited by another type, tested with an ANOVA analysis at a key electrode and within the time window of interest. All evidence was consistent with the hypothesis that semantic, grammar and harmonic processes rely, at least to an extent, on separable neurocognitive substrates. This section summarises the evidence and demonstrates the replicability of these ERP effects across experiments.

The N400 effects were characterised by a centroparietal increase in negative amplitude for semantic word violations compared to standard words roughly between 300–450 ms, see Figure 6.1. These characteristics are similar to those presented in previous studies (Gutierrez et al., 2012; Kamp et al., 2015; Kutas & Federmeier, 2010; Kutas & Hillyard, 1983; Moreno & Vázquez, 2011; Nigam et al., 1992; Rommers et al., 2013; Tiedt et al., 2020). The N400 was not elicited by grammar or harmonic expectancy violations, and it had a distinct centroparietal scalp topography. The evidence is consistent with three hypotheses. First, semantic processing relies on separable mechanisms from grammar processing, as proposed in the Faculty of Language model proposed by Hauser et al. (2002). Second, the current findings are consistent with previous evidence that finds no N400 for musical stimuli (Besson & Faïta, 1995; Besson & Macar, 1987; Miranda & Ullman, 2007; Paller et al., 1992), and models suggesting separable neurocognitive substrates for semantic processing between language and music domains (Brown et al., 2006). Third, there were no interactions in the N400 when it was presented with regular or irregular meter, suggesting separable mechanisms for semantic and meter processing. Overall, the current results support that the N400 indicates neurocognitive substrates that are specialised for semantic processing.

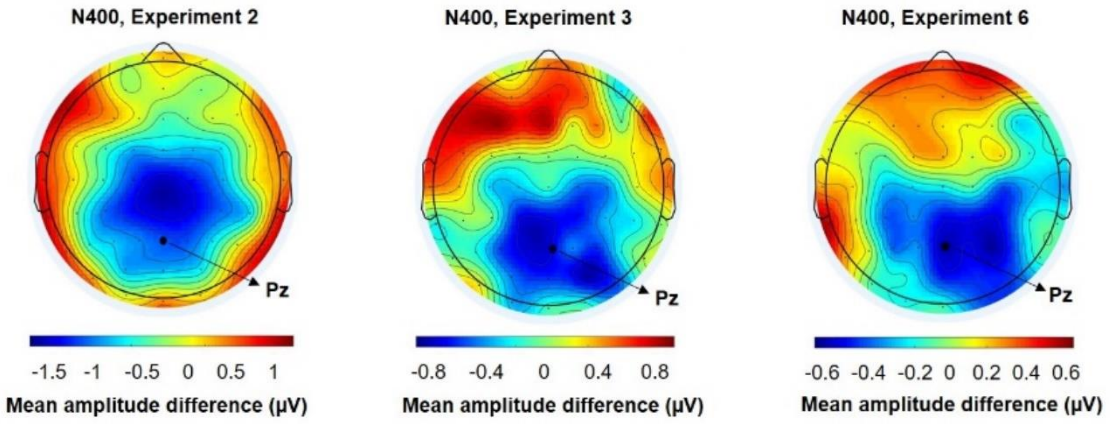


Figure 6.1 – Replication of the N400 semantic expectancy violation effect across experiments. Scalp topography maps of the mean amplitude difference (µV) between 300–450 ms. The mean amplitude difference was calculated as the mean amplitude (µV) of the unrelated semantic violations minus the standard words. Please note: these scalp maps are on different scales.

A P600 effect was elicited by grammar expectancy violations in Experiments 2, 3 and 6, consistent with those presented previous studies (Hahne et al., 2012; Mehravari et al., 2015; Osterhout & Holcomb, 1992). It was characterised by a central increase in positivity for grammar violations compared to standard words between 500–800 ms, see Figure 6.2. Crucially, the P600 was not elicited by semantic or harmonic expectancy violations, and it had a distinct central scalp topography. The P600 findings are consistent with three hypotheses. First, grammar processing relies on separable mechanisms from semantic processing, which is consistent with the Faculty of Language model and previous empirical findings (Hauser et al., 2002; Kim & Osterhout, 2005). Second, grammar processing relies, at least to an extent, on separable neurocognitive substrates from harmony processing (Frisch et al., 2003; Osterhout, 1999). Third, as there were no interactions in the P600 when it was presented with regular or irregular meter, the current findings support the hypothesis that there are separate neurocognitive substrates that are specialised for grammar and meter processing. Overall, the current results support the existence of specialised neurocognitive substrates for grammar processing.

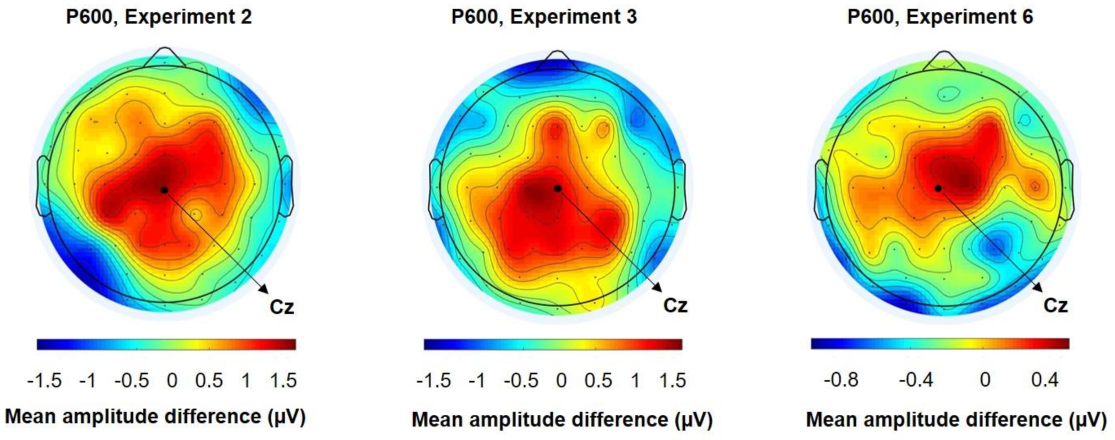


Figure 6.2 – Replication of the P600 grammar expectancy violation effect across experiments. Scalp topography maps of the mean amplitude difference (μV) between 500–800 ms after the final (target) word's onset. The mean amplitude difference was calculated as the mean amplitude (μV) of the grammar expectancy violation minus the standard words. Please note: these scalp maps are on different scales.

Harmonic expectancy violations elicited a P300 effect, as in previous studies (Beisteiner et al., 1999; Carrión & Bly, 2008; Janata, 1995; Loehr et al., 2013). This effect was characterised by a frontal increase in positive amplitude roughly between 250–400 ms after the final (target) word's onset, see Figure 6.3. The P300 was not elicited by semantic or grammar expectancy violations and had a distinct frontal scalp topography. The P300 evidence is consistent with three hypotheses. First, harmony processing relies on separable neurocognitive substrates from semantic processes (Besson & Faïta, 1995; Besson & Macar, 1987; Miranda & Ullman, 2007; Paller et al., 1992). Second, it relies on separable neurocognitive substrates from grammar processes (Frisch et al., 2003; Osterhout, 1999). Third, the P300 effect was not affected when harmonic violations were presented simultaneously with meter violations. The model that Peretz and Colthart (2003) proposed for separable mechanisms for meter and pitch processing might extend from pitch to some aspects of harmony processing. Overall, the P300 appears to indicate specialised neurocognitive substrates for harmony processing.

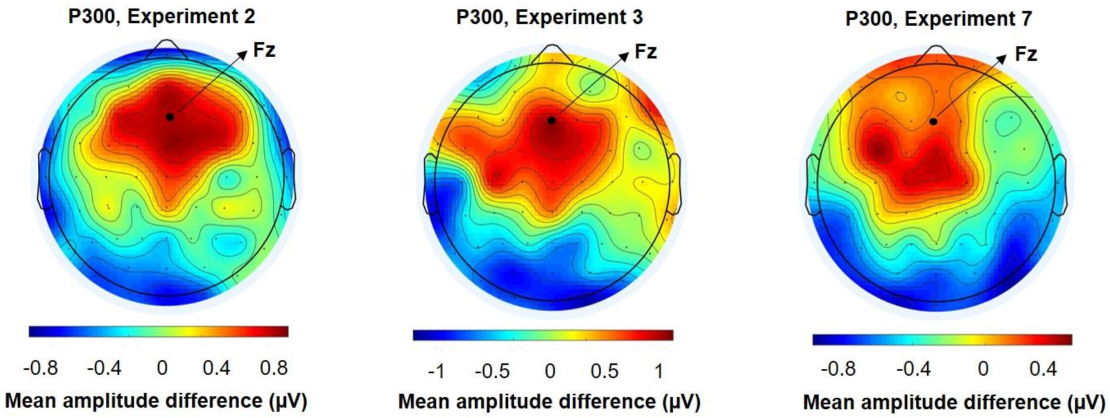


Figure 6.3 – Replication of the P300 harmonic expectancy violation effect across experiments. Scalp topography maps of the mean amplitude difference (μV) between 250–400 ms after the final (target) chord's onset. The mean amplitude difference (μV) was calculated as the mean amplitude (μV) of the unrelated harmony chords minus the tonic (I) chords. Please note: these scalp maps are on different scales.

P600 and P300 comparison

Despite compelling arguments that there are links between neurocognitive substrates underlying grammar and harmony processing, and links between P600 and P300 effects within the language domain (Frisch et al., 2003; Osterhout, 1999; Sassenhagen & Fiebach, 2019; Tabullo et al., 2013), no previous studies appear to have actively compared P600 effects for grammar expectancy violations with P300 effects for harmonic expectancy violations.

There are two findings in the current experiments that suggest that the P600 for grammar expectancy violations and the P300 for harmonic expectancy violations do not indicate the same neurocognitive substrates. First, the two ERP effects have different scalp topographies – the P300 has a frontal scalp distribution, and the P600 has a central scalp distribution. Therefore, the idea that these two ERP effects indicate the same neurocognitive substrates just at a different time point, as has been suggested for P300 and P600 effects within the language domain (Sassenhagen & Fiebach, 2019), is not supported. Second, in the current experiments, a P300 was not elicited by grammar expectancy violations, and harmonic expectancy violations did not elicit a P600. This finding is unlike a study by Patel et al. (1998), who report a P600 effect for harmonic violations. However, as discussed in section 1.3.2, it is unclear whether the “P600 effect” presented by Patel (1998) is a P600 effect, as it occurred 1100 ms after the onset of the harmonic violation. To aid the debate, when discussing the P600/P300 relationship in future research, it would be beneficial to refer to the effects in more detail, perhaps including details of the scalp topography, and with what stimuli their temporal characteristics refer to – for example, “a frontal P300, peaking at 352 ms after the onset of harmonic expectancy violations”. The current results suggest that the P600 indicates neurocognitive substrates that are specialised for grammar processing and the P300 indicates substrates that are specialised for harmonic processing.

6.1.4. A domain-general ERP effect

This thesis identified a potentially domain-general ERP effect (the P2) for meter expectancy violations in language and music domains. The P2 meter violation effect was significant in both language and music tasks across different samples of participants and within- and between-subjects experimental designs. It was present when the meter violations were presented on their own (Experiment 5) and when they were presented simultaneously with semantic, grammar and harmonic expectancy violations (Experiments 6 and 7), see Figure 6.4.

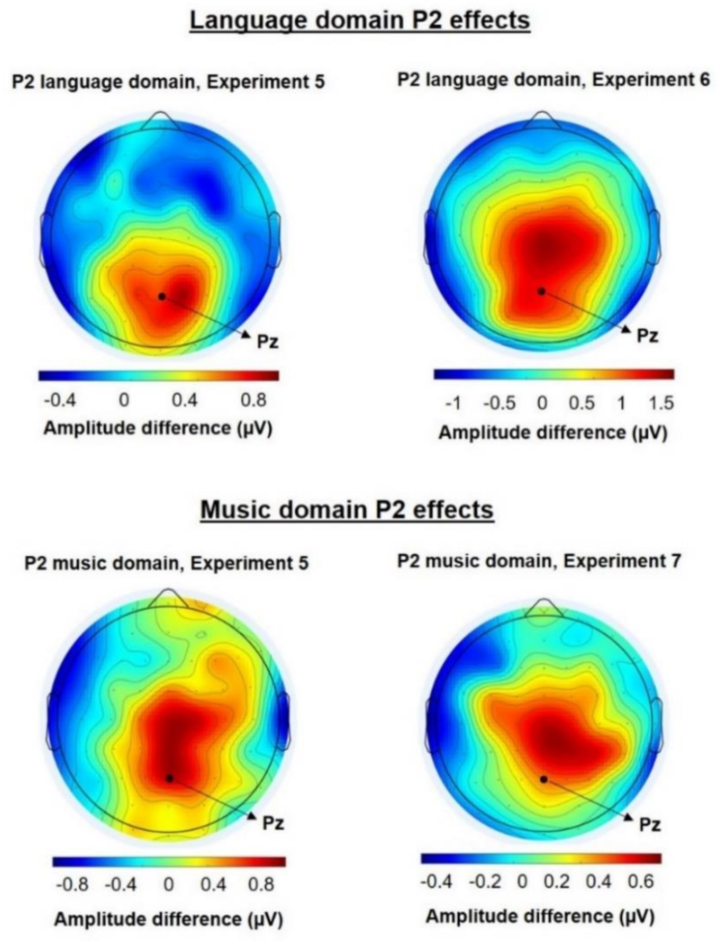


Figure 6.4 – The P2 effects across experiments in language and music domains (-200 to 0 ms baselines). Scalp topography maps of the amplitude difference (µV) at 200 ms. The amplitude difference (µV) was calculated as the amplitude (µV) of the irregular 500/1500 ms interstimulus intervals minus the regular (1000 ms) interstimulus intervals. Please note: these scalp maps are on different scales.

The P2 effects are characterised by an increase in positivity for stimuli presented at an irregular interstimulus interval (short 500 ms and long 1500 ms) compared to a regular interstimulus interval (1000 ms). These P2 effects are consistent with previous studies that report P2 effects for interstimulus interval violations in the music domain (Neuhaus & Knösche, 2008; Raij et al., 1997). However, previous studies have not investigated interstimulus interval violations in the language domain, despite an underlying sense of regular meter playing a potentially significant role in language comprehension (Cutler & Otake, 2002; Pitt & Samuel, 1990; Port, 2003). P2 meter violation effects have not been previously reported for meter violations in the language domain. The occurrence of P2 effects in both language and music tasks provides original support for the hypothesis that there are domain-general neurocognitive substrates for meter processing. These domain-general mechanisms might have evolved to aid the efficient processing of upcoming events, by directing attention to specific time points, as proposed by the Dynamic Attending Theory (Jones & Boltz, 1989). This interpretation is supported, as P2 effects have previously been linked to temporal attention orientation (Liu et al., 2013).

There were no statistical interactions in the P2 effects between language and music tasks in the ANOVA analyses, supporting the hypothesis of domain-general meter processing neurocognitive substrates. The P2 in the language domain in Experiment 5 (top left of Figure 6.4) appears to have a more posterior scalp distribution than the other effects. However, the P2 in the language domain in Experiment 6 appears to have similar scalp topography to the P2 effects in the music domain. EEG has a poor spatial resolution, limiting interpretation of the neural sources of these effects. This issue is known as a spatial inverse problem, and it is a general limitation of the ERP method (Poldrack, 2006; Tarantola, 2004). Due to fMRI's high spatial resolution (Huster et al., 2012), combining EEG and fMRI methods could shed further light on the spatial characteristics of P2 effects for meter violations. Consequently, this approach could further test the hypothesis that these P2 effects indicate shared neurocognitive substrates for meter processing in language and music domains.

There were no interactions in the P2 effects for meter violations, and the N400 effect for semantic violations, P600 effect for grammar violations or P300 effect for harmony violations. Similar to how Peretz and Colthart (2003) suggested that meter processing relies on separable mechanisms compared to pitch processing, the current findings suggest that neurocognitive substrates involved in meter processing might be separable from those involved in semantic, grammar and harmony processing, at least in terms of those indicated by the current ERP effects. Overall, the current evidence suggests that the P2 effects could indicate specialised mechanisms for meter processing that are domain-general.

6.1.5. An unclassified ERP effect

There was clear evidence that an N1 effect was elicited by meter violations in the music domain (see Figure 6.5). It was found in both Experiment 5 (when meter violations were presented on their own) and Experiment 7 (when meter violations were presented simultaneously with harmonic expectancy violations). This N1 effect was characterised by a frontal increase in negativity for chords presented at irregular interstimulus intervals compared to regular interstimulus intervals and was measured at Fz between 80–140 ms (the a-priori N1 time window). Similar N1 effects have been reported for different types of meter violations in the music domain in previous studies (Fitzroy & Sanders, 2015; Habibi et al., 2014; Kober et al., 2015). Therefore, the current evidence support that the N1 effect is a reliable effect for meter violations in the music domain. There were no interactions in the N1 or P300 effects when meter violations were presented simultaneously with harmony violations. Therefore, the current findings support the hypothesis that, at least to some extent, there is neurocognitive separation for meter and harmony processing. The model that Peretz and Coltheart (2003) proposed for separations between meter and pitch processing might extend to some types of harmony processing.

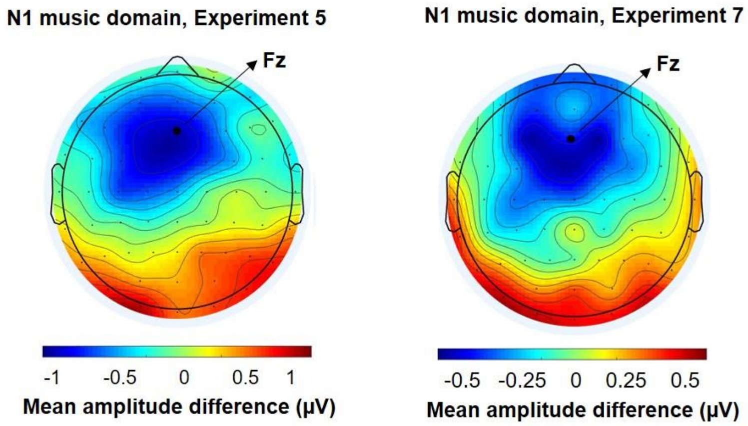


Figure 6.5 – N1 music domain meter violation effect graphs showing replication across experiments (-200 to 0 ms baselines). Scalp topography maps of the mean amplitude difference (µV) between 80–140 ms after the final (target) chord's onset. The mean amplitude difference (µV) was calculated as the mean amplitude (µV) of the irregular 500/1500 ms minus the regular (1000 ms) interstimulus intervals. Please note: these scalp maps are on different scales.

On the other hand, the evidence for an N1 effect for meter violations in the language domain was inconclusive. The results of Experiment 5 showed a main effect of mean amplitude for the interstimulus intervals in the a-priori N1 time window (80–140 ms) at Fz with the –200 to 0 ms baseline, but pairwise comparisons showed that this was only significant for the short and not for the long interstimulus intervals. Additionally, there was no significant N1 effect in the 0 to +50 ms baseline analysis. Later experiments combined the short 500 ms and long 1500 ms interstimulus intervals to make an “irregular interstimulus interval” condition. Therefore, it was deemed inappropriate for the later experiments to investigate an N1 effect for meter violations in the language domain, as any significant effects of “irregular interstimulus interval” might actually be due to the short 500 ms interstimulus interval, and not both short 500 ms and long 1500 ms interstimulus intervals.

Overall, in the current experiments, the tests of N1 meter violation effects do not provide strong support for either domain-specialised or domain-general neurocognitive substrates for meter processing. However, previous research reports N1 effects for meter violations in the language domain (Cox et al., 2016; Zhang & Zhang, 2019). Further research is needed to understand the N1 effects of meter violations in the language domain. To extend the current experiments, interstimulus interval violations could be presented in the auditory modality in addition to the visual modality in the language domain, to investigate whether the N1 effect is more pronounced for auditory stimuli than visual stimuli. If so, this could explain why the N1 was not as clear in the language domain compared to the music domain, and why it was not significant, when it has been reported in previous studies of meter violations in the language domain (which have tended to present stimuli in the auditory modality). Moreover, in previous studies that report an N1 effect for meter violations in the language domain, irregular patterns of stressed and non-stressed syllables (iambic or trochaic) were used to create meter violations (Cox et al., 2016; Zhang & Zhang, 2019). This idea could be extended to compare ERP effects between language and music domains by presenting similar patterns of stressed and non-stressed beats in language and music stimuli.

6.1.6. An absence of expertise effects

This thesis also aimed to investigate whether there was any evidence of cross-domain effects of English language or musical expertise on ERP effects. Such effects may have provided evidence for shared neurocognitive substrates between language and music domains, as proposed by the OPERA hypothesis (Patel, 2011). There was no evidence of cross-domain effects of either English language or musical training expertise, for any ERP effects. However, the lack of expertise effects may be due to the current sampling methods, which might not have been adequate to test expertise effects.

Participants were not recruited based on English language experience, in any of the experiments. For analysis, they were split into two groups: “experts” were native English speakers and “non-experts” were non-native English speakers. All participants attended university at an English-speaking university, and English proficiency levels were very high for non-native English speakers in all experiments. Of the 116 participants who filled in the English Language Questionnaire, 71 were native English speakers, and 45 were non-native English speakers. When participants rated their English language proficiency on a scale of 1–10 (see Appendix D), the ratings of both groups were quite high (Native speakers: *M* = 9.97, *SD* = 0.24; Non-native speakers: *M* = 8.60, *SD* = 1.18). Therefore, it is likely that there was not enough difference in English proficiency to detect English language expertise effects. Future research aiming to investigate differences in ERP effects based on English language experience should aim to recruit a non-native speaker group with lower proficiency levels than the current experiments.

Similarly, for the EEG experiments, participants were not recruited based on musical training experience, as heightening participants' awareness of their musical ability before the ERP tasks could have had additional effects that may have affected ERP measurements (Luck, 2014). Based on the median of accumulated practice time across experiments, the cut-off chosen for amateur musicians was > 1000 hours of musical training, and for non-musicians, < 1000 hours of musical training (see Appendix F). This may not have been a distinct enough cut off to test the effects of musical training robustly. Future studies with more distinct, and perhaps a-priori, cut-offs for the groups of amateur musicians and non-musicians could be valuable for testing the moderating effects of musical training.

Crucially, due to these limitations of the methods, the lack of expertise effects in the current experiments neither provides evidence for or against either domain-specialised or domain-general neurocognitive substrates for processing in language and music domains.

6.2. General limitations and future research directions

After additional task-related components were found for the explicit music task in Experiment 2, and not for the implicit task in Experiment 3, all EEG experiments adopted implicit tasks. This method also allowed for different types of expectancy violations to be presented simultaneously in Experiments 6 and 7, with the task not creating a bias towards one type of expectancy violation. However, visual comparisons of the ERP effects between Experiments 2 and 3 suggested that the ERP effects for expectancy violations were not as clear for implicit tasks compared to explicit tasks. This reduced clarity could be due to the task-driven saliency of the expectancy violations. Alternatively, it could be due to participants' general attention to the stimuli. In the explicit task, participants were required to answer whether the final word or chord was "expected?" or "unexpected?". Therefore, attention was drawn to the final word or chord by the task. To successfully complete the implicit tasks, it was not necessary for participants to read the final word (they could just look at the colour of the sentences – green or blue), or to pay attention to the final chord (they were asked to focus on the instrument playing the chord progressions – piano or organ). This issue was deemed likely to be more prominent in the language task than the music task, as it was easier for participants not to read the words of the sentences than to stop hearing the chords that were playing over the loudspeakers. With this in mind, to encourage the participants to pay attention to the whole sentences, they were told that there would be questions to answer about the content of the sentences after the presentation of the experimental trials, as has been done in previous studies (Goregliad Fjaellingsdal et al., 2016; Kamp et al., 2015; Kutas & Hillyard, 1983). Across the five questions that were asked, participants' response accuracy was *M* = 75% (*SD* = 21%) in Experiment 3, *M* = 74% (*SD* = 31%) in Experiment 5 and *M* = 74% (*SD* = 27%) in Experiment 6. Considering that the questions were about the sentence stems and these were presented for all conditions (so each was presented three or more times in each experiment), these accuracy numbers seem quite low. Future studies could adopt additional methods to measure participants' attention to the stimuli, such as an eye-tracking device, to be more confident that participants are reading the sentences in implicit tasks.

The pre-stimulus effect for the interstimulus interval manipulations remains unexplained. This effect occurred for all interstimulus interval manipulations, in Experiments 5, 6 and 7. The cluster-based permutation analyses were limited to an epoch between 0 to 800 ms after target onset. The significant clusters started at, or close to, 0 ms. Previous studies suggest that meter processing effects can be pre-attentive, so can be early-onset (Bouwer et al., 2014; Silva & Castro, 2019). However, in the current experiments, the effect in question is difficult to explain, as it appears to have occurred before the onset of the final word or chord. In Experiments 6 and 7, there was no pre-stimulus effect for the semantic, grammar or harmonic analyses, but there was for the interstimulus interval analyses. Therefore, the pre-stimulus effect does not seem to be a preprocessing issue, but something to do with the nature of the meter manipulations. Future research, specifically designed to investigate this pre-stimulus effect is needed. For such research, all stimuli that build anticipation of a target stimulus should be included in analysis. This approach is recommended because the pre-stimulus effect might be a kind of contingent negative variation (CNV) effect, which is a slow negative wave that builds with the anticipation of the occurrence of a stimulus (Kononowicz & Penney, 2016).

The current thesis's results support that ERP correlates of expectancy violations can provide valuable insights into the debate about shared or separable neurocognitive substrates between domains. With this in mind, there are four proposed direct extensions of the experiments presented in this thesis. First, as the P2 is present in language and music domains and visual and auditory modalities, the extent of this effect's generalisability should be investigated further. For example, an experiment could test whether presenting interstimulus interval based meter violations presented through other modalities such as tactile stimulation (e.g., small vibrations to participants' hands) elicits a similar P2 effect. A study found that when participants' finger tapping was aligned with the meter of heard sentences, detection of word changes was enhanced (Falk & Dalla Bella, 2016). This finding suggests that meter processing mechanisms might extend to the touch modality, and increases the motivation to investigate whether a P2 effect similar to those found in the current experiments occurs for meter violations in the touch modality. This would shed further light on whether the neurocognitive substrates indicated by the P2 effect are specific to meter violations in language and music domains, or whether it is an ERP correlate of meter violations in general.

Second, future research could present semantic and grammar expectancy violations simultaneously with harmonic expectancy violations. This direction is facilitated by presenting the language stimuli in the visual modality, and the music stimuli in the auditory modality, allowing for interactions in ERP effects between domains to be tested. If, for example, future research finds interactions in the simultaneous grammar and harmonic expectancy violations but not in the simultaneous semantic and harmonic expectancy violations, there could be novel support for the hypothesis for shared neurocognitive substrates for grammar and harmony processing in the P600 and P300 effects (Bornkessel-Schlesewsky et al., 2011; Coulson et al., 1998; Münte et al., 1998; Sassenhagen et al., 2014). Alternatively, if there are no interactions in these effects for simultaneously presented violation types, the case for domain-specialised processes could be strengthened.

Third, future research could investigate whether P300 effects for harmonic violations, such as those presented in this thesis, are modality-dependent or modality-independent. Professional musicians train to be able to read music and follow harmonic progressions visually (Gudmundsdottir, 2010). Therefore, one approach to investigate this question could be to ask professional musicians to read sheet music, in which there are congruent and incongruent harmonic stimuli. One might predict that there would be an increase in amplitude positivity at around 300 ms after participants see the incongruent harmonic stimuli compared to the congruent harmonic stimuli (a P300 effect). In a within-subjects design, the same stimuli could be presented in both visual and auditory modalities, to investigate modality-dependent differences in P300 characteristics. Such experiments would further test whether P300 effects elicited by harmonic violations represent neurocognitive substrates specific to cognitive (opposed to just sensory) harmonic processing. In turn, this could lead to stronger conclusions about domain-specialised neurocognitive mechanisms in language and music domains and could increase the flexibility of future research designs to include both (or either) visual and auditory presentations of harmonic expectancy violations. This type of flexibility is currently present in semantic and grammar expectancy violation research, as previous research reports that the N400 effect for semantic violations and P600 effect for grammar expectancy violations are identical when presented in visual and auditory modalities (Balconi & Pozzoli, 2005; Kutas & Federmeier, 2000).

Fourth, artificial intelligence methods could give additional insights into the distinctness of the ERP effects found in the current thesis. Machine learning models, such as multivariate pattern analysis, could be trained to classify the ERP effects, as in Sassenhagen and Fiebach (2019). For example, a classifier could be trained to distinguish between regular and irregular interstimulus intervals in the music domain. If this classifier performed equally well at distinguishing between regular and irregular interstimulus intervals in the language domain, as for the music domain, then the case for domain-general neurocognitive substrates for meter processing could be supported.

Lastly, it is worth restating that a likely highly beneficial general direction for future research would be to investigate the spatial characteristics of the current ERP effects further. For example, by combining EEG and fMRI methods, it could be possible to shed further light on the distinctness of the neurocognitive substrates that are indicated by the ERP effects. This approach would help to overcome the low spatial resolution issue with the EEG methodology, which results in a spatial inverse problem (Tarantola, 2004), and could provide valuable insights into the question of shared and separable neurocognitive substrates between language and music domains. Source analysis with EEG data is an alternative option, although there are still debates about the best methods to use for source analysis, see Awan et al. (2019) for a recent review. To attempt source analysis, it is generally recommended to use a scalp cap with more electrodes than was used in the current experiments and to take accurate scalp measurements of each participant. The scalp caps used in the current experiments had 64 electrodes, and the recommended number for source analysis tends to be between 128–256 electrodes (Michel & Brunet, 2019).

6.3. Practical implications

The P2 effect reported in the current thesis supports the hypothesis that there are domain-general neurocognitive substrates for meter processing. Therefore, it supports the idea that meter training could be emphasised in interventions aiming to improve abilities in each domain, as previously suggested by some other authors (Overy et al., 2003; Schon & Tillmann, 2015). For example, a study found that school students who practised synchronising movements to a metronome beat had more improved reading fluency scores than a control group, who attended regular school lessons during that time (Taub & Lazarus, 2013).

Furthermore, if future research is consistent with the hypothesis that meter processing relies on domain-general mechanisms, then this could have a ripple effect for interpretations of studies and links between language and music domain. For example, a study of healthy adults showed a clear association between prosody perception and music perception, but the authors narrowed this down to being specific to meter perception in the music domain (Hausen et al., 2013). The authors suggested that there was a cross-domain expertise effect. However, it is possible that findings such as this one could be better explained by the existence of domain-general neurocognitive substrates that operate in both language and music domains.

The findings of this thesis could contribute to further applications. For example, with further testing, the expectancy violations investigated in the current thesis could aid our understanding of the neurocognitive correlates of dyslexia, which are elusive (Linkersdorfer, 2011). People with dyslexia have been found to have poorer meter skills (Corriveau & Goswami, 2008; Richardson et al., 2004; Sallat & Jentschke, 2015; Wolff, 2002). Therefore, it could be valuable for future research to investigate whether meter violation ERP effects differ between people with and without dyslexia. Similar research could shed light on other disorders too. For example, studies could test whether people with congenital amusia have a reduced or absent P300 harmonic violation effect. Following this approach could aid understanding of several disorders while also shedding more light on the extent of shared and separable neurocognitive substrates between language and music domains.

Finally, recent research suggests that a reduced amplitude of the N400 effect for semantic expectancy violations is a potential risk factor for pathological cognitive ageing (Paitel et al., 2021), and Alzheimer’s disease (Joyal et al., 2020). The authors of these papers suggest that the N400 might be able to predict cognitive decline in adults who are currently healthy but at risk of pathological cognitive decline. Equipment needed to measure ERPs is cheaper than other neuroimaging methods. Therefore, it is possible that measurements of the N400 could be clinically relevant and could be a useful addition to the assessment of risk factors for pathological cognitive ageing (Joyal et al., 2020; Paitel et al., 2021). Perhaps the addition of the ERP effects for grammar (P600), harmony (P300) and meter (N1 and P2) violations could improve the accuracy of such assessments. The current thesis suggests that ERP effects for these different types of expectancy violations indicate different neurocognitive substrates and, consequently, it is possible that their inclusion could lead to more accurate predictions of pathological cognitive decline. This area of emerging research could transform the implications and applications of ERP research.

6.4. **Conclusion**

To conclude, this thesis provides valuable insights into the question of which processes might rely on domain-specialised neurocognitive substrates and which might rely on domain-general ones in language and music domains. The experiments provide evidence to support the distinctness of ERP effects associated with semantic, grammar and harmonic expectancy violations. Respectively, these are a centroparietal N400 effect, a central P600 effect and a frontal P300 effect. This evidence supports the hypothesis that domain-specialised neurocognitive substrates are involved in these processes. Additionally, ERP correlates of meter violations were actively compared between language and music domains for the first time, and a potentially domain-general central P2 effect was identified. This P2 effect provides novel support for the hypothesis that meter processing relies, at least to an extent, on domain-general neurocognitive substrates. Lastly, there were no interactions in ERP effects when meter violations were presented simultaneously with semantic, grammar and harmony violations. Therefore, the hypothesis that these ERP effects rely on process-specialied neurocognitive mechanisms is supported. It is hoped that this thesis might motivate further interdisciplinary investigations, exploring the extent of shared and separable neurocognitive substrates between language and music domains.

References

Antovic, M. (2009). Towards the Semantics of Music: The Twentieth Century. *Language & History, 52*, 119-129. https://doi.org/10.1179/175975309X452003

Anurova, I., & Immonen, P. (2017). Native vs. second language discrimination in the modulation of the N400/N400m. *European Journal of Neuroscience, 45*(10), 1289-1299. https://doi.org/10.1111/ejn.13566

Ashley, R. (2002). Do[[n't]] Change a Hair for Me: The Art of Jazz Rubato. *Music Perception, 19*(3), 311-332. https://doi.org/10.1525/mp.2002.19.3.311

Awan, F. G., Saleem, O., & Kiran, A. (2019). Recent trends and advances in solving the inverse problem for EEG source localization. *Inverse Problems in Science and Engineering, 27*(11), 1521-1536. https://doi.org/10.1080/17415977.2018.1490279

Bach-Gesellschaft. (1892). Johann Sebastian Bach's Werke. On *Bach-Gesellschaft Ausgabe, (Vol. 39)*. Breitkopf und Härtel.

Balconi, M., & Pozzoli, U. (2005). Comprehending Semantic and Grammatical Violations in Italian. N400 and P600 Comparison with Visual and Auditory Stimuli. *Journal of psycholinguistic research, 34*, 71-98. https://doi.org/10.1007/s10936-005-3633-6

Batterink, L., & Neville, H. (2013). Implicit and explicit second language training recruit common neural mechanisms for syntactic processing. *Journal of cognitive neuroscience, 25*(6), 936-951. https://doi.org/10.1162/jocn_a_00354

Beisteiner, R., Erdler, M., Mayer, D., Gartus, A., Edward, V., Kaindl, T., Golaszewski, S., Lindinger, G., & Deecke, L. (1999). A marker for differentiation of capabilities for processing of musical harmonies as detected by magnetoencephalography in musicians. *Neuroscience letters, 277*(1), 37-40. https://doi.org/10.1016/s0304-3940(99)00836-8

Benjamin, W. E. (1984). A Theory of Musical Meter. *Music Perception, 1*(4), 355-413. https://doi.org/10.2307/40285269

Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21*(6), 1278-1296. https://doi.org/10.1037/0096-1523.21.6.1278

Besson, M., Faita, F., Czternasty, C., & Kutas, M. (1997). What's in a pause: event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology, 46*(1), 3-23. https://doi.org/https://doi.org/10.1016/S0301-0511(96)05215-5

Besson, M., Faïta, F., Peretz, I., Bonnel, A. M., & Requin, J. (1998). Singing in the Brain: Independence of Lyrics and Tunes. *Psychological Science, 9*(6), 494-498. http://www.jstor.org/stable/40063362

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology, 24*(1), 14-25. https://doi.org/10.1111/j.1469-8986.1987.tb01853.x

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100*(1), 100-130. https://doi.org/10.1016/j.cognition.2005.11.007

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods, 42*(3), 665-670. https://doi.org/10.3758/BRM.42.3.665

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition, 8*(6), 631-642. https://doi.org/10.3758/BF03213783

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language, 117*(3), 133-152. https://doi.org/doi: 10.1016/j.bandl.2010.09.010

Bouwer, F. L., Van Zuijen, T. L., & Honing, H. (2014). Beat Processing Is Pre-Attentive for Metrically Simple Rhythms with Clear Accents: An ERP Study [Article]. *PLoS One, 9*(5), 9, Article e97467. https://doi.org/10.1371/journal.pone.0097467

Breen, M., Fitzroy, A. B., & Ali, M. O. (2019). Event-Related Potential Evidence of Implicit Metric Structure during Silent Reading. *Brain Sciences, 9*(8), 28, Article 192. https://doi.org/10.3390/brainsci9080192

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience, 23*(10), 2791-2803. https://doi.org/https://doi.org/10.1111/j.1460-9568.2006.04785.x

Canette, L. H., Fiveash, A., Krzonowski, J., Corneyllie, A., Lalitte, P., Thompson, D., Trainor, L., Bedoin, N., & Tillmann, B. (2020). Regular rhythmic primes boost P600 in grammatical error processing in dyslexic adults and matched controls. *Neuropsychologia, 138*, 12, Article 107324. https://doi.org/10.1016/j.neuropsychologia.2019.107324

Carrión, R. E., & Bly, B. M. (2008). The effects of learning on event-related potential correlates of musical expectancy. *Psychophysiology, 45*(5), 759-775. https://doi.org/https://doi.org/10.1111/j.1469-8986.2008.00687.x

Cason, N., Astesano, C., & Schon, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica, 155*, 43-50. https://doi.org/10.1016/j.actpsy.2014.12.002

Chee, M. W. L., O'Craven, K. M., Bergida, R., Rosen, B. R., & Savoy, R. L. (1999). Auditory and visual word processing studied with fMRI. *Human brain mapping, 7*(1), 15-28. https://doi.org/https://doi.org/10.1002/(SICI)1097-0193(1999)7:1

Chern, A., Tillmann, B., Vaughan, C., & Gordon, R. L. (2018). New evidence of a rhythmic priming effect that enhances grammaticality judgments in children. *Journal of Experimental Child Psychology, 173*, 371-379. https://doi.org/10.1016/j.jecp.2018.04.007

Cheung, V. K. M., Meyer, L., Friederici, A. D., & Koelsch, S. (2018). The right inferior frontal gyrus processes nested non-local dependencies in music. *Scientific Reports, 8*, 12, Article 3822. https://doi.org/10.1038/s41598-018-22144-9

Coco, M. I., Araujo, S., & Petersson, K. M. (2017). Disentangling stimulus plausibility and contextual congruency. *Electro-physiological evidence for differential cognitive dynamics, 96*, 150-163. https://doi.org/10.1016/j.neuropsychologia.2016.12.008

Corriveau, K., & Goswami, U. (2008). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex, 45*, 119-130. https://doi.org/10.1016/j.cortex.2007.09.008

Coulson, S., King, J., & Kutas, M. (1998). Expect the Unexpected: Event-related Brain Response to Morphosyntactic Violations. *Language and Cognitive Processes, 13*, 21-58. https://doi.org/10.1080/016909698386582

Cox, S. R., Dickie, D. A., Ritchie, S. J., Karama, S., Pattie, A., Royle, N. A., Corley, J., Aribisala, B. S., Valdés Hernández, M., Muñoz Maniega, S., Starr, J. M., Bastin, M. E., Evans, A. C., Wardlaw, J. M., & Deary, I. J. (2016). Associations between education and brain structure at age 73 years, adjusted for age 11 IQ. *Neurology, 87*(17), 1820-1826. https://doi.org/10.1212/WNL.0000000000003247

Cuskley, C., Colaiori, F., Castellano, C., Loreto, V., Pugliese, M., & Tria, F. (2015). The adoption of linguistic rules in native and non-native speakers: Evidence from a Wug task. *Journal of Memory and Language, 84*, 205-223. https://doi.org/https://doi.org/10.1016/j.jml.2015.06.005

Cutler, A., & Otake, T. (2002, 2002/02/01/). Rhythmic Categories in Spoken-Word Recognition. *Journal of Memory and Language, 46*(2), 296-322. https://doi.org/https://doi.org/10.1006/jmla.2001.2814

Davis, H., Mast, T., Yoshie, N., & Zerlin, S. (1966). The slow response of the human cortex to auditory stimuli: Recovery process. *Electroencephalography and Clinical Neurophysiology, 21*(2), 105-113. https://doi.org/https://doi.org/10.1016/0013-4694(66)90118-0

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast brain plasticity during word learning in musically-trained children. *Frontiers in Human Neuroscience, 11*. https://doi.org/https://doi.org/10.3389/fnhum.2017.00233

Ellison, D., Moisseinen, N., Fachner, J., & Brattico, E. (2015). Affective versus cognitive responses to musical chords: An ERP and behavioral study. *Psychomusicology: Music, Mind, and Brain, 25*, 423-434. https://doi.org/10.1037/pmu0000127

Elmer, S., Meyer, M., & Jancke, L. (2010). Simultaneous interpreters as a model for neuronal adaptation in the domain of language processing. *Brain Research, 1317*, 147-156. https://doi.org/10.1016/j.brainres.2009.12.052

Erlbeck, H., Kübler, A., Kotchoubey, B., & Veser, S. (2014). Task instructions modulate the attentional mode affecting the auditory MMN and the semantic N400.

Frontiers in Human Neuroscience, 8, 654-654. https://doi.org/10.3389/fnhum.2014.00654

Falk, S., & Dalla Bella, S. (2016). It is better when expected: aligning speech and motor rhythms enhances verbal processing. *Language Cognition and Neuroscience, 31*(5), 699-708. https://doi.org/10.1080/23273798.2016.1144892

Faroqi-Shah, Y., Slevc, L. R., Saxena, S., Fisher, S. J., & Pifer, M. (2020). Relationship between musical and language abilities in post-stroke aphasia. *Aphasiology, 34*(7), 793-819. https://doi.org/10.1080/02687038.2019.1650159

Fields, E. C. (2017). Factorial Mass Univariate ERP Toolbox [Computer software]. https://github.com/ericcfields/FMUT/releases

Fields, E. C., & Kuperberg, G. R. (2020). Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data. *Psychophysiology, 57*(2), e13468. https://doi.org/10.1111/psyp.13468

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry, N. W., Jr. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology, 20*(4), 400-409. https://doi.org/10.1111/j.1469-8986.1983.tb00920.x

Fitzroy, A. B., & Sanders, L. D. (2015). Musical Meter Modulates the Allocation of Attention across Time. *Journal of cognitive neuroscience, 27*(12), 2339-2351. https://doi.org/10.1162/jocn_a_00862

Foldal, M. D., Blenkmann, A. O., Llorens, A., Knight, R. T., Solbakk, A. K., & Endestad, T. (2020). The brain tracks auditory rhythm predictability independent of selective attention. *Scientific Reports, 10*(1), 13, Article 7975. https://doi.org/10.1038/s41598-020-64758-y

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews, 25*(4), 355-373.

Frisch, S., Kotz, S. A., von Cramon, D. Y., & Friederici, A. D. (2003). Why the P600 is not just a P300: The role of the basal ganglia. *Clinical Neurophysiology, 114*(2), 336-340.

Gao, J., Tian, H., Yang, Y., Yu, X., Li, C., & Rao, N. (2014). A novel algorithm to enhance P300 in single trials: application to lie detection using F-score and SVM. *PLoS One, 9*(11), e109700-e109700. https://doi.org/10.1371/journal.pone.0109700

George, E. M., & Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia, 49*(5), 1083-1094. https://doi.org/10.1016/j.neuropsychologia.2011.02.001

Goldman, A., Jackson, T., & Sajda, P. (2018). Improvisation experience predicts how musicians categorize musical structures. *Psychology of Music, 48*(1), 18-34. https://doi.org/10.1177/0305735618779444

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. https://doi.org/https://doi.org/10.1111/nyas.12683

Goregliad Fjaellingsdal, T., Ruigendijk, E., Scherbaum, S., & Bleichner, M. G. (2016). The N400 Effect during Speaker-Switch—Towards a Conversational Approach of Measuring Neural Correlates of Language. *Frontiers in psychology, 7*, 1854. https://doi.org/10.3389/fpsyg.2016.01854

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25*(2), 149-188. https://doi.org/10.1080/01690960902965951

Groppe, D. M., Makeig, S., & Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *NeuroImage, 45*(4), 1199-1211. https://doi.org/10.1016/j.neuroimage.2008.12.038

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology, 48*(12), 1711-1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Gudmundsdottir, H. R. (2010). Advances in music-reading research. *Music Education Research, 12*(4), 331-338. https://doi.org/10.1080/14613808.2010.504809

Gunter, T. C., 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. https://doi.org/10.1162/089892900562336

Guo, S., & Koelsch, S. (2016). Effects of veridical expectations on syntax processing in music: Event-related potential evidence. *Scientific Reports, 6*(1), 19064. https://doi.org/10.1038/srep19064

Gutierrez, E., Williams, D., Grosvald, M., & Corina, D. (2012). Lexical access in American Sign Language: an ERP investigation of effects of semantics and phonology. *Brain Res, 1468*, 63-83. https://doi.org/10.1016/j.brainres.2012.04.029

Habibi, A., Wirantana, V., & Starr, A. (2014). Cortical activity during perception of musical rhythm: Comparing musicians and nonmusicians [doi:10.1037/pmu0000046]. *Psychomusicology, 24*(2), 125-135. https://doi.org/10.1037/pmu0000046

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13*(3), 339-356. https://doi.org/https://doi.org/10.1016/S0926-6410(01)00127-6

Hahne, A., Wolf, A., Müller, J., Mürbe, D., & Friederici, A. (2012). Sentence comprehension in proficient adult cochlear implant users: On the vulnerability of syntax. *Language and Cognitive Processes, 27*, 1-13. https://doi.org/10.1080/01690965.2011.653251

Haupt, F. S., Schlesewsky, M., Roehm, D., Friederici, A. D., & Bornkessel-Schlesewsky, I. (2008). The status of subject–object reanalyses in the language comprehension architecture. *Journal of Memory and Language, 59*(1), 54-96. https://doi.org/https://doi.org/10.1016/j.jml.2008.02.003

Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Sarkamo, T. (2013). Music and speech prosody: a common rhythm. *Frontiers in psychology, 4*, 16, Article 566. https://doi.org/10.3389/fpsyg.2013.00566

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science, 298*(5598), 1569-1579. https://doi.org/10.1126/science.298.5598.1569

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402. https://doi.org/10.1038/nrn2113

Hofman, S., Klein, C., & Arlazoroff, A. (1993). Common hemisphericity of language and music in a musician: A case report. *Journal of Communication Disorders, 26*(2), 73-82. https://doi.org/10.1016/0021-9924(93)90001-Q

Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex, 47*(6), 674-689. https://doi.org/10.1016/j.cortex.2010.07.010

Huster, R. J., Debener, S., Eichele, T., & Herrmann, C. S. (2012). Methods for Simultaneous EEG-fMRI: An Introductory Review. *The Journal of Neuroscience, 32*(18), 6053-6060. https://doi.org/10.1523/jneurosci.0447-12.2012

Jackendoff, R. (2009). Parallels and Nonparallels between Language and Music. *Music Perception, 26*, 195-204. https://doi.org/10.1525/mp.2009.26.3.195

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: what is it, and what's special about it? *Cognition, 100*(1), 33-72. https://doi.org/10.1016/j.cognition.2005.11.005

James, C. E., Britz, J., Vuilleumier, P., Hauert, C.-A., & Michel, C. M. (2008). Early neuronal responses in right limbic structures mediate harmony incongruity processing in musical experts. *NeuroImage, 42*(4), 1597-1608. https://doi.org/https://doi.org/10.1016/j.neuroimage.2008.06.025

James, C. E., Oechslin, M. S., Michel, C. M., & De Pretto, M. (2017). Electrical Neuroimaging of Music Processing Reveals Mid-Latency Changes with Level of Musical Expertise. *Frontiers in neuroscience, 11*, 613. https://doi.org/10.3389/fnins.2017.00613

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of cognitive neuroscience, 7*(2), 153-164. https://doi.org/10.1162/jocn.1995.7.2.153

Janata, P., Tillman, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 121-140. https://doi.org/10.3758/CABN.2.2.121

Jaśkiewicz, M., Francuz, P., Zabielska-Mendyk, E., Zapała, D., & Augustynowicz, P. (2016). Effects of harmonics on aesthetic judgments of music: an ERP study involving laypersons and experts. *Acta Neurobiologiae Experimentalis, 76*(2), 142-151. https://doi.org/10.21307/ane-2017-013

Jentzsch, I., Mkrtchian, A., & Kansal, N. (2014). Improved effectiveness of performance monitoring in amateur instrumental musicians. *Neuropsychologia, 52*(100), 117-124. https://doi.org/10.1016/j.neuropsychologia.2013.09.025

Johnson, M. L., & Larson, S. (2003). "Something in the Way She Moves"-Metaphors of Musical Motion. *Metaphor and Symbol, 18*(2), 63-84. https://doi.org/10.1207/S15327868MS1802_1

Jonaitis, E. M., & Saffran, J. R. (2009). Learning Harmony: The Role of Serial Statistics. *Cognitive Science, 33*(5), 951-968. https://doi.org/10.1111/j.1551-6709.2009.01036.x

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychol Rev, 96*(3), 459-491. https://doi.org/10.1037/0033-295x.96.3.459

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol Sci, 13*(4), 313-319. https://doi.org/10.1111/1467-9280.00458

Jongsma, M. L. A., Eichele, T., Quiroga, R. Q., Jenks, K. M., Desain, P., Honing, H., & Van Rijn, C. M. (2005). Expectancy effects on omission evoked potentials in musicians and non-musicians. *Psychophysiology, 42*(2), 191-201. https://doi.org/10.1111/j.1469-8986.2005.00269.x

Joyal, M., Groleau, C., Bouchard, C., Wilson, M. A., & Fecteau, S. (2020). Semantic Processing in Healthy Aging and Alzheimer's Disease: A Systematic Review of the N400 Differences. *Brain Sciences, 10*(11), 40, Article 770. https://doi.org/10.3390/brainsci10110770

Jung, H., Sontag, S., Park, Y. S., & Loui, P. (2015). Rhythmic Effects of Syntax Processing in Music and Language. *Frontiers in psychology, 6*, 11, Article 1762. https://doi.org/10.3389/fpsyg.2015.01762

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cogn Psychol, 39*(3-4), 159-207. https://doi.org/10.1006/cogp.1999.0716

Kalender, B., Trehub, S. E., & Schellenberg, E. G. (2013). Cross-cultural differences in meter perception. *Psychological Research, 77*, 196-203. https://doi.org/10.1007/s00426-012-0427-y

Kamp, S. M., Potts, G. F., & Donchin, E. (2015). On the roles of distinctiveness and semantic expectancies in episodic encoding of emotional words. *Psychophysiology, 52*(12), 1599-1609. https://doi.org/10.1111/psyp.12537

Kappenman, E., & Luck, S. (2012). *The Oxford Handbook of Event-Related Potential Components*. https://doi.org/10.1093/oxfordhb/9780195374148.001.0001

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205-225. https://doi.org/10.1016/j.jml.2004.10.002

Kim, S.-G., Kim, J. S., & Chung, C. K. (2011). The Effect of Conditional Probability of Chord Progression on Brain Response: An MEG Study. *PLoS One, 6*(2), e17337. https://doi.org/10.1371/journal.pone.0017337

Kisser, J. E., Wendell, C. R., Spencer, R. J., & Waldstein, S. R. (2012). Neuropsychological Performance of Native versus Non-native English Speakers. *Archives of Clinical Neuropsychology, 27*(7), 749-755. https://doi.org/10.1093/arclin/acs082

Kober, S. E., Schweiger, D., Witte, M., Reichert, J. L., Grieshofer, P., Neuper, C., & Wood, G. (2015). Specific effects of EEG based neurofeedback training on memory functions in post-stroke victims. *Journal of NeuroEngineering and Rehabilitation, 12*(1), 107. https://doi.org/10.1186/s12984-015-0105-6

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

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: an fMRI study. *NeuroImage, 25*(4), 1068-1076. https://doi.org/10.1016/j.neuroimage.2004.12.050

Kolk, H., & Chwilla, D. (2007). Late Positivities in unusual situations. *Brain and Language, 100*, 257-261. https://doi.org/10.1016/j.bandl.2006.07.006

Kononowicz, T. W., & Penney, T. B. (2016). The contingent negative variation (CNV): timing isn’t everything. *Current Opinion in Behavioral Sciences, 8*, 231-237. https://doi.org/https://doi.org/10.1016/j.cobeha.2016.02.022

Krohn, K., Brattico, E., Välimäki, V., & Tervaniemi, M. (2007). Neural Representations Of The Hierarchical Sale Pitch Structure. *Music Perception, 24*(3), 281-296.

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), 16, Article e0141069. https://doi.org/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), 150685. https://doi.org/doi:10.1098/rsos.150685

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 4*(12), 463-470. https://doi.org/10.1016/s1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2010). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual review of psychology, 62*(1), 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language, 11*(2), 354-373. https://doi.org/https://doi.org/10.1016/0093-934X(80)90133-9

Kutas, M., & Hillyard, S. A. (1983). Event-Related Brain Potentials to Grammatical Errors and Semantic Anomalies. *Memory & Cognition, 11*(5), 539-550. https://doi.org/Doi 10.3758/Bf03196991

Lemhofer, K., Schrieters, H., & Indefrey, P. (2020). Syntactic Processing in L2 Depends on Perceived Reliability of the Input: Evidence From P600 Responses to Correct Input. *Journal of Experimental Psychology-Learning Memory and Cognition, 46*(10), 1948-1965. https://doi.org/10.1037/xlm0000895

Lense, M. D., & Dykens, E. M. (2016). Beat Perception and Sociability: Evidence from Williams Syndrome. *Frontiers in psychology, 7*, 13, Article 886. https://doi.org/10.3389/fpsyg.2016.00886

Leong, V., & Goswami, U. (2014). Assessment of rhythmic entrainment at multiple timescales in dyslexia: Evidence for disruption to syllable timing. *Hearing Research, 308*, 141-161. https://doi.org/10.1016/j.heares.2013.07.015

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

Li, X. Q., Shao, X. M., Xia, J. Y., & Xu, X. Y. (2019). The cognitive and neural oscillatory mechanisms underlying the facilitating effect of rhythm regularity on speech comprehension. *Journal of Neurolinguistics, 49*, 155-167. https://doi.org/10.1016/j.jneuroling.2018.05.004

Li, Y., Wang, Y., Zhang, B., Wang, Y., & Zhou, X. (2018). Electrophysiological Responses to Expectancy Violations in Semantic and Gambling Tasks: A

Comparison of Different EEG Reference Approaches. *Frontiers in neuroscience, 12*, 169. https://doi.org/10.3389/fnins.2018.00169

Linkersdorfer, J. (2011). Neurocognitive correlates of dyslexia. *Kindheit Und Entwicklung, 20*(1), 4-12. https://doi.org/10.1026/0942-5403/a000035

Liu, Y., Zhang, D., Ma, J., Li, D., Yin, H., & Luo, Y. (2013). The Attention Modulation on Timing: An Event-Related Potential Study. *PLoS One, 8*(6), e66190. https://doi.org/10.1371/journal.pone.0066190

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring Individual and Joint Action Outcomes in Duet Music Performance. *Journal of cognitive neuroscience, 25*(7), 1049-1061. https://doi.org/10.1162/jocn_a_00388

Loerts, H., Stowe, L. A., & Schmid, M. S. (2013). Predictability speeds up the re-analysis process: An ERP investigation of gender agreement and cloze probability. *Journal of Neurolinguistics, 26*(5), 561-580. https://doi.org/https://doi.org/10.1016/j.jneuroling.2013.03.003

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*(213). https://doi.org/10.3389/fnhum.2014.00213

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique: Second Edition*. MIT Press. ISBN: 978-0-262-52585-5.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146-157. https://doi.org/10.1111/psyp.12639

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience, 4*(5), 540-545. https://doi.org/10.1038/87502

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of Syllabic Lengthening on Semantic Processing in Spoken French: Behavioral and Electrophysiological Evidence. *Cerebral Cortex, 17*(11), 2659-2668. https://doi.org/10.1093/cercor/bhl174

Magne, C., Jordan, D. K., & Gordon, R. L. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain and Language, 153-154*, 13-19. https://doi.org/10.1016/j.bandl.2016.01.001

Mattys, S. L., & Samuel, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language, 36*(1), 87-116. https://doi.org/10.1006/jmla.1996.2472

McNeill, D. (1994). Hand and Mind: What Gestures Reveal About Thought. *Bibliovault OAI Repository, the University of Chicago Press, 27*. https://doi.org/10.2307/1576015

Mehravari, A. S., Tanner, D., Wampler, E. K., Valentine, G. D., & Osterhout, L. (2015). Effects of Grammaticality and Morphological Complexity on the P600 Event-Related Potential Component. *PLoS One, 10*(10), e0140850. https://doi.org/10.1371/journal.pone.0140850

Menceloglu, M., Grabowecky, M., & Suzuki, S. (2020). Rhythm Violation Enhances Auditory-Evoked Responses to the Extent of Overriding Sensory Adaptation in Passive Listening. *Journal of cognitive neuroscience, 32*(9), 1654-1671. https://doi.org/10.1162/jocn_a_01578

Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago Press. ISBN: 9780226521381.

Michel, C. M., & Brunet, D. (2019). EEG Source Imaging: A Practical Review of the Analysis Steps. *Frontiers in neurology, 10*, 325-325. https://doi.org/10.3389/fneur.2019.00325

Midorikawa, A., Kawamura, M., & Kezuka, M. (2003). Musical Alexia for Rhythm Notation: A Discrepancy Between Pitch and Rhythm. *Neurocase, 9*(3), 232-238. https://doi.org/10.1076/neur.9.3.232.15558

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. https://doi.org/10.1016/j.neuroimage.2007.07.034

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology, 48*(2), 229-240. https://doi.org/10.1111/j.1469-8986.2010.01061.x

Moreno, E. M., Casado, P., & Martín-Loeches, M. (2016). Tell me sweet little lies: An event-related potentials study on the processing of social lies. *Cognitive, Affective, & Behavioral Neuroscience, 16*(4), 616-625. https://doi.org/10.3758/s13415-016-0418-3

Moreno, E. M., & Vázquez, C. (2011). Will the glass be half full or half empty? Brain potentials and emotional expectations. *Biological Psychology, 88*(1), 131-140. https://doi.org/https://doi.org/10.1016/j.biopsycho.2011.07.003

Morey, R. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*. https://doi.org/10.20982/tqmp.04.2.p061

Müller, M., Höfel, L., Brattico, E., & Jacobsen, T. (2010). Aesthetic judgments of music in experts and laypersons--an ERP study. *International Journal of Psychophysiology, 76*(1), 40-51. https://doi.org/10.1016/j.ijpsycho.2010.02.002

Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*(3), 217-226. https://doi.org/https://doi.org/10.1016/S0028-3932(97)00119-X

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. . In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 463-472). MIT Press. ISBN: 9780262232067.

Neuhaus, C., & Knösche, T. R. (2008). Processing of pitch and time sequences in music. *Neurosci Lett, 441*(1), 11-15. https://doi.org/10.1016/j.neulet.2008.05.101

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience, 14*(9), 1105-1107. https://doi.org/10.1038/nn.2886

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of cognitive neuroscience, 4*(1), 15-22. https://doi.org/10.1162/jocn.1992.4.1.15

Núñez-Peña, M. I., & Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research, 18*(2), 130-141. https://doi.org/10.1016/j.cogbrainres.2003.09.010

Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., & Kotz, S. (2013). Aesthetic and Emotional Effects of Meter and Rhyme in Poetry. *Frontiers in psychology, 4*(10). https://doi.org/10.3389/fpsyg.2013.00010

Ochoa, C. J., & Polich, J. (2000). P300 and blink instructions. *Clinical Neurophysiology, 111*(1), 93-98. https://doi.org/10.1016/s1388-2457(99)00209-6

Osterhout, L. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14*(1), 1-14. https://doi.org/https://doi.org/10.1080/016909699386356

Osterhout, L., Allen, M. D., McLaughlin, J., & Inoue, K. (2002). Brain potentials elicited by prose-embedded linguistic anomalies. *Memory & Cognition, 30*(8), 1304-1312. https://doi.org/10.3758/BF03213412

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785-806. https://doi.org/https://doi.org/10.1016/0749-596X(92)90039-Z

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of cognitive neuroscience, 8*(6), 507-526. https://doi.org/10.1162/jocn.1996.8.6.507

Overy, K., Nicolson, R. I., Fawcett, A. J., & Clarke, E. F. (2003). Dyslexia and music: Measuring musical timing skills. *Dyslexia, 9*(1), 18-36. https://doi.org/10.1002/dys.233

Paitel, E. R., Samii, M. R., & Nielson, K. A. (2021). A systematic review of cognitive event-related potentials in mild cognitive impairment and Alzheimer's disease. *Behavioural Brain Research, 396*, 17, Article 112904. https://doi.org/10.1016/j.bbr.2020.112904

Paller, K. A., McCarthy, G., & Wood, C. C. (1992). Event-related potentials elicited by deviant endings to melodies. *Psychophysiology, 29*(2), 202-206. https://doi.org/10.1111/j.1469-8986.1992.tb01686.x

Palmer, B., Nasman, V. T., & Wilson, G. F. (1994). Task decision difficulty: effects on ERPs in a same-different letter classification task. *Biological Psychology, 38*(2), 199-214. https://doi.org/https://doi.org/10.1016/0301-0511(94)90039-6

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology - Human Perception and Performance, 16*, 728-741. https://doi.org/10.1037/0096-1523.16.4.728

Palolahti, M., Leino, S., Jokela, M., Kopra, K., & Paavilainen, P. (2005). Event-related potentials suggest early interaction between syntax and semantics during on-line sentence comprehension. *Neuroscience letters, 384*, 222-227. https://doi.org/10.1016/j.neulet.2005.04.076

Patel, A. (2008). *Music, Language, and the Brain*. Oxford University Press. ISBN: 9780195123753.

Patel, A., & Daniele, J. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87*, 35-45. https://doi.org/10.1016/S0010-0277(02)00187-7

Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in psychology, 2*, 142-142. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of cognitive neuroscience, 10*(6), 717-733. https://doi.org/10.1162/089892998563121

Patel, A. D., Iversen, J. R., Wassenaar, M., & Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology, 22*(7-8), 776-789. https://doi.org/10.1080/02687030701803804

Pereira, D., Cardoso, S., Ferreira-Santos, F., Fernandes, C., Cunha-Reis, C., Paiva, T., Almeida, P., Silveira, C., Barbosa, F., & Teixeira, J. (2014). Effects of Inter-stimulus Interval (ISI) Duration on the N1 and P2 Components of the Auditory Event-Related Potential. *International Journal of Psychophysiology, doi*. https://doi.org/10.1016/j.ijpsycho.2014.09.012

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nat Neurosci, 6*(7), 688-691. https://doi.org/10.1038/nn1083

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Science, 7*(8), 362-367. https://doi.org/10.1016/s1364-6613(03)00150-5

Peretz, I., & Zatorre, R. (2005). Brain Organization for Music Processing. *Annual review of psychology, 56*, 89-114. https://doi.org/10.1146/annurev.psych.56.091103.070225

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. https://doi.org/10.3758/s13423-012-0344-5

Phillips-Silver, J., Toiviainen, P., Gosselin, N., & Peretz, I. (2013). Amusic does not mean unmusical: beat perception and synchronization ability despite pitch deafness. *Cognitive Neuropsychology, 30*(5), 311-331. https://doi.org/10.1080/02643294.2013.863183

Picton, L., Saunders, B., & Jentzsch, I. (2012). "I will fix only my own mistakes": An ERP study investigating error processing in a joint choice-RT task. *Neuropsychologia, 50*(5), 777-785. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2012.01.011

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 564-573. https://doi.org/10.1037//0096-1523.16.3.564

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Science, 10*(2), 59-63. https://doi.org/10.1016/j.tics.2005.12.004

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J., & Donchin, E. (1988). P300 and the word frequency effect. *Electroencephalography and Clinical Neurophysiology, 70*(1), 33-45. https://doi.org/https://doi.org/10.1016/0013-4694(88)90192-7

Politimou, N., Dalla Bella, S., Farrugia, N., & Franco, F. (2019). Born to Speak and Sing: Musical Predictors of Language Development in Pre-schoolers. *Frontiers in psychology, 10*, 18, Article 948. https://doi.org/10.3389/fpsyg.2019.00948

Port, R. F. (2003). Meter and speech. *Journal of Phonetics, 31*(3-4), 599-611. https://doi.org/10.1016/j.wocn.2003.08.001

Poulin-Charronnat, B., Bigand, E., & Koelsch, S. (2006). Processing of musical syntax tonic versus subdominant: an event-related potential study. *Journal of cognitive neuroscience, 18*(9), 1545-1554. https://doi.org/10.1162/jocn.2006.18.9.1545

Raij, T., McEvoy, L., Mäkelä, J. P., & Hari, R. (1997). Human auditory cortex is activated by omissions of auditory stimuli. *Brain Research, 745*(1), 134-143. https://doi.org/https://doi.org/10.1016/S0006-8993(96)01140-7

Relander, K., Rämä, P., & Kujala, T. (2009). Word Semantics Is Processed Even without Attentional Effort. *Journal of cognitive neuroscience, 21*(8), 1511-1522. https://doi.org/10.1162/jocn.2009.21127

Richardson, U., Thomson, J. M., Scott, S. K., & Goswami, U. (2004). Auditory processing skills and phonological representation in Dyslexic children. *Dyslexia, 10*(3), 215-233. https://doi.org/https://doi.org/10.1002/dys.276

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Rohrmeier, M., & Cross, I. (2008). Statistical Properties of Tonal Harmony in Bach's Chorales. *Proceedings of the 10th International Conference on Music Perception and Cognition*.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of cognitive neuroscience, 25*(5), 762-776. https://doi.org/10.1162/jocn_a_00337

Rosslau, K., Herholz, S. C., Knief, A., Ortmann, M., Deuster, D., Schmidt, C. M., Zehnhoff-Dinnesen, A., Pantev, C., & Dobel, C. (2016). Song Perception by Professional Singers and Actors: An MEG Study. *PLoS One, 11*(2), e0147986. https://doi.org/10.1371/journal.pone.0147986

Rothermich, K., Schmidt-Kassow, M., Schwartze, M., & Kotz, S. A. (2010). Event-related potential responses to metric violations: Rules versus meaning. *NeuroReport: For Rapid Communication of Neuroscience Research, 21*(8), 580-584. https://doi.org/10.1097/WNR.0b013e32833a7da7

Sallat, S., & Jentschke, S. (2015). Music Perception Influences Language Acquisition: Melodic and Rhythmic-Melodic Perception in Children with Specific Language Impairment. *Behavioural Neurology, 2015*, 606470. https://doi.org/10.1155/2015/606470

Sarfarazi, M., Cave, B., Richardson, A., Behan, J., & Sedgwick, E. (1999). Visual event related potentials modulated by contextually relevant and irrelevant olfactory primes. *Chemical Senses, 24*, 145-154. https://doi.org/https://doi.org/10.1093/chemse/24.2.145

Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage, 200*, 425-436. https://doi.org/https://doi.org/10.1016/j.neuroimage.2019.06.048

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language, 137*, 29-39. https://doi.org/https://doi.org/10.1016/j.bandl.2014.07.010

Schmidt-Kassow, M., & Kotz, S. A. (2008). Entrainment of syntactic processing? ERP-responses to predictable time intervals during syntactic reanalysis. *Brain Research, 1226*, 144-155. https://doi.org/10.1016/j.brainres.2008.06.017

Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of cognitive neuroscience, 21*(9), 1693-1708. https://doi.org/10.1162/jocn.2008.21153

Schmuckler, M. A., & Boltz, M. G. (1994). Harmonic and rhythmic influences on musical expectancy. *Perception & psychophysics, 56*(3), 313-325. https://doi.org/10.3758/BF03209765

Schon, D., & Tillmann, B. (2015). Short- and long-term rhythmic interventions: perspectives for language rehabilitation. In E. Bigand, B. Tillmann, I. Peretz, R. J. Zatorre, L. Lopez, & M. Majno (Eds.), *Neurosciences and Music V: Cognitive Stimulation and Rehabilitation* (Vol. 1337, pp. 32-39). Blackwell Science Publishers. https://doi.org/10.1111/nyas.12635

Seger, C. A., Spiering, B. J., Sares, A. G., Quraini, S. I., Alpeter, C., David, J., & Thaut, M. H. (2013). Corticostriatal contributions to musical expectancy perception. *Journal of cognitive neuroscience, 25*(7), 1062-1077. https://doi.org/10.1162/jocn_a_00371

Shelburne, S. A. (1972). Visual evoked responses to word and nonsense syllable stimuli. *Electroencephalography and Clinical Neurophysiology, 32*(1), 17-25. https://doi.org/https://doi.org/10.1016/0013-4694(72)90224-6

Silva, S., & Castro, S. (2019). Structural meter perception is pre-attentive. *Neuropsychologia, 133*, 12, Article 107184. https://doi.org/10.1016/j.neuropsychologia.2019.107184

Slater, J., Tierney, A., & Kraus, N. (2013). At-Risk Elementary School Children with One Year of Classroom Music Instruction Are Better at Keeping a Beat. *PLoS One, 8*(10), 9, Article e77250. https://doi.org/10.1371/journal.pone.0077250

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. https://doi.org/10.3758/16.2.374

Sloboda, J. A., & Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude. *Music Perception, 19*(1), 87-120. https://doi.org/10.1525/mp.2001.19.1.87

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *NeuroImage, 17*(3), 1613-1622. https://doi.org/10.1006/nimg.2002.1288

Spencer, A. (1995). *Phonology: Theory and description* (Vol. 9). Wiley-Blackwell. ISBN: 0631192336.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. *Journal of cognitive neuroscience, 18*(8), 1380-1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Suppanen, E., Huotilainen, M., & Ylinen, S. (2019). Rhythmic structure facilitates learning from auditory input in newborn infants. *Infant Behaviour and Development, 57*, 101346. https://doi.org/10.1016/j.infbeh.2019.101346

Tabullo, A., Sevilla, Y., Segura, E., Zanutto, S., & Wainselboim, A. (2013). An ERP study of structural anomalies in native and semantic free artificial grammar: Evidence for shared processing mechanisms. *Brain Research, 1527*, 149-160. https://doi.org/10.1016/j.brainres.2013.05.022

Tarantola, A. (2004). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics. ISBN: 0898715725.

Taub, G., & Lazarus, P. (2013). The Effects Of Training In Timing And Rhythm On Reading Achievement. *Issues in Education, 5*, 343-350. https://doi.org/10.19030/cier.v5i4.7598

Thomson, J. M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of research in reading, 29*(3), 334-348. https://doi.org/https://doi.org/10.1111/j.1467-9817.2006.00312.x

Tiedt, H. O., Ehlen, F., & Klostermann, F. (2020). Age-related dissociation of N400 effect and lexical priming. *Scientific Reports, 10*(1), 20291. https://doi.org/10.1038/s41598-020-77116-9

Tierney, A. T., & Kraus, N. (2013). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain and Language, 124*(3), 225-231. https://doi.org/https://doi.org/10.1016/j.bandl.2012.12.014

Tillmann, B., Koelsch, S., Escoffier, N., Bigand, E., Lalitte, P., Friederici, A. D., & von Cramon, D. Y. (2006). Cognitive priming in sung and instrumental music: activation of inferior frontal cortex. *NeuroImage, 31*(4), 1771-1782. https://doi.org/10.1016/j.neuroimage.2006.02.028

Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research, 70*(5), 345-358. https://doi.org/10.1007/s00426-005-0222-0

Trainor, L., & Trehub, S. (1994). Key membership and implied harmony in Western tonal music: developmental perspectives. *Perception & psychophysics, 56*, 125-132. https://doi.org/https://doi.org/10.3758/BF03213891

Trenkic, D., Mirkovic, J., & Altmann, G. T. M. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition, 17*(2), 237-257. https://doi.org/10.1017/S1366728913000321

van de Meerendonk, N., Indefrey, P., Chwilla, D. J., & Kolk, H. H. J. (2011). Monitoring in language perception: Electrophysiological and hemodynamic responses to spelling violations. *NeuroImage, 54*(3), 2350-2363. https://doi.org/https://doi.org/10.1016/j.neuroimage.2010.10.022

Vignolo, L. A. (2003). Music Agnosia and Auditory Agnosia. *Annals of the New York Academy of Sciences, 999*(1), 50-57. https://doi.org/10.1196/annals.1284.005

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Ostergaard, L. (2005). To musicians, the message is in the meter - Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *NeuroImage, 24*(2), 560-564. https://doi.org/10.1016/j.neuroimage.2004.08.039

Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Frontiers in psychology, 5*(1111). https://doi.org/10.3389/fpsyg.2014.01111

Warrick, M. J., Kibler, A. W., & Topmiller, D. A. (1965). Response time to unexpected stimuli. *Human Factors,* *7*(1), 81-86. https://doi.org/10.1177/001872086500700110

Wicha, N., Moreno, E., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of cognitive neuroscience, 16*, 1272-1288. https://doi.org/10.1162/0898929041920487

Wickham, H. (2009). ggplot2: elegant graphics for data analysis. *Springer New York, http://had.co.nz/ggplot2/book*.

Wolff, P. H. (2002). Timing precision and rhythm in developmental dyslexia. *Reading and Writing, 15*(1), 179-206. https://doi.org/10.1023/A:1013880723925

Yates, C. M., Justus, T., Atalay, N. B., Mert, N., & Trehub, S. E. (2017). Effects of musical training and culture on meter perception. *Psychology of Music, 45*(2), 231-245. https://doi.org/10.1177/0305735616657407

Zhang, J. J., Che, X. C., & Yang, Y. F. (2019). Event-related brain potentials suggest a late interaction of pitch and time in music perception. *Neuropsychologia, 132*, 7, Article Unsp 107118. https://doi.org/10.1016/j.neuropsychologia.2019.107118

Zhang, N., & Zhang, Q. F. (2019). Rhythmic pattern facilitates speech production: An ERP study. *Scientific Reports,* *9*, 11, Article 12974. https://doi.org/10.1038/s41598-019-49375-8

Zhao, T. C., Lam, H. T. G., Sohi, H., & Kuhl, P. K. (2017). Neural processing of musical meter in musicians and non-musicians. *Neuropsychologia, 106*, 289-297. https://doi.org/10.1016/j.neuropsychologia.2017.10.007

Zioga, I., Di Bernardi Luft, C., & Bhattacharya, J. (2016). Musical training shapes neural responses to melodic and prosodic expectation. *Brain Research, 1650*, 267-282. https://doi.org/10.1016/j.brainres.2016.09.015

Appendices

Appendix A. 24 sentences for language stimuli

Table A1 – The 24 sentence stems word, related semantic violation, unrelated semantic violation, and grammar expectancy violation, adapted from Block and Baldwin (2010).

Sentence stem	Standard word	Related semantic	Unrelated semantic	Grammar expectancy violation
The cleaner dusted the books on the	shelf.	shrine.	salad.	shelved.
Kate put the flowers in an expensive	vase.	car.	fork.	vases.
He washed the dirty dishes in the	sink.	bathtub.	pencil.	sinking.
The athlete enjoyed lifting weights at a	gym.	shop.	compass.	gyms.
At night the elderly woman locks a	door.	museum.	spoon.	doors.
Surfers are scared of getting bitten by	sharks.	camels.	rice.	shark.
Spring was Mark's favourite season of the	year.	flowers.	phone.	yeared.
When babies are hungry they will often	cry.	crawl.	bike.	cries.
He cashed his new paycheck at the	bank.	school.	elephant.	banked.
George must keep his dog on a	leash.	table.	cup.	leashed.
The small toddler could already say three	words.	hymns.	chairs.	worded.
She graduated at the top of her	class.	stage.	paintbrush.	classed.
In the quiet movie theatre Kim's phone	rang.	sang.	walked.	ringed.
Bob proposed and gave her a diamond	ring.	tiara.	biscuit.	rings.
The wonderful waitress got a generous	tip.	party.	bin.	tipped.
The princess would someday become a	queen.	king.	mushroom.	queenly.
The wealthy child attended a private	school.	house.	comb.	schooling.
The lecture should last about one	hour.	week.	stick.	hourly.
John swept the floor with a	broom.	toothbrush.	computer.	brooms.
Water and sunshine help plants to	grow.	climb.	throw.	grew.
Father carved the turkey with a	knife.	saw.	tissue.	knives.
Her new shoes were the wrong	size.	height.	peak.	sizely.
In the first space enter your	name.	price.	hair.	named.
He loosened the tie around his	neck.	finger.	cloud.	necking.

Appendix B. Music stimuli, progression origins

Table B1 – Details of the 72 chord progressions used in Experiment 1. For all following Experiments (2, 3, 4, 5 and 7), only chord progression stems with the "related end types" of iii and IV were used, making up the 24 chord progressions. These 24 chord progressions are shaded in grey. The chorales were taken from a collection of Bach's chorales – Bach-Gesellschaft (1892).

Bach's chorale name	Chorale number	Bars	Experiment key	Related end type
Herr, ich denk an jene Zeit	76	12 to 14	C	ii
Gott lebet noch	67	36 to 38	C#	ii
Gottlob, es geht nunmehr zu Ende	68	14 to 16	D	ii
Gott sei gelobet und gebenedeiet	69	18 to 20	Eb	ii
Heilig, heilig	72	16 to 17	E	ii
Herr Gott, dich loben wir	73	13 to 16	F	ii
Vor deinen Thron tret ich hiermit	74	13 to 16	F#	ii
Herr, wie du willst, so schick's mit mir	86	11 to 13	G	ii
Herzlich lieb hab ich dich, o Herr	87	17 to 19	Ab	ii
Hilf, Gott, daß mirs gelinge	90	9 to 13	Ab	ii
Ich dank dir, Gott, für all Wohltat	93	12 to 15	Bb	ii
Ich dank dir, lieber Herre	94	2 to 4	B	ii
Ach bleib bei uns, Herr Jesu Christ	1	8 to 10	C	iii
Ach Gott und Herr	3	6 to 8	C#	iii
Der Tag, der ist so freudenreich	41	15 to 16	D	iii
Dir, dir, Jehova, will ich singen	46	14 to 16	Eb	iii
Alles ist an Gottes Segen	11	11 to 12	E	iii
Als der gütige Gott	12	8 to 10	F	iii
Eins ist not, ach Herr, dies Eine	51	22 to 24	F#	iii
Erstanden ist der heilge Christ	53	14 to 16	G	iii
Aus meines Herzens Grunde	17	18 to 21	Ab	iii
Christus, der ist mein Leben	28	6 to 8	Ab	iii
Gott der Vater wohn uns bei	64	15 to 16	Bb	iii
Gottes Sohn ist kommen	65	12 to 13	B	iii
Jesu, meines Herzens Freud	108	11 to 13	C	IV
Jesu, nun sei gepreiset	109	7 to 9	C#	IV
Jesu, nun sei gepreiset	109	26 to 30	D	IV
Jesus Christus, unser Heiland	111	9 to 11	Eb	IV
Jesus, meine Zuversicht	112	8 to 9	E	IV
In dulci jubilo	115	12 to 16	F	IV
Komm, Gott Schöpfer, Heiliger Geist	117	1 to 2	F#	IV
Kyrie, Gott Vater in Ewigkeit	118	31 to 33	G	IV
Lass, o Herr, dein Ohr sich neigen	119	10 to 12	Ab	IV
Lobt Gott, ihr Christen, allzugleich	122	2 to 4	Ab	IV
Lobt Gott, ihr Christen, allzugleich	122	8 to 10	Bb	IV

Lobt Gott, ihr Christen, allzugleich	123	8 to 10	B	IV
Machs mit mir, Gott, nach deiner Güt	124	6 to 8	C	V
Mein Augen schließ ich jetzt	125	2 to 4	C#	V
Meinen Jesum laß ich nicht	127	12 to 13	D	V
Meines Lebens letzte Zeit	128	3 to 4	Eb	V
Mitten wir im Leben sind	130	6 to 8	E	V
Nicht so traurig, nicht so sehr	131	2 to 4	F	V
Nun freut euch, lieben Christen g'mein	135	2 to 4	F#	V
Nun freut euch, lieben Christen g'mein	135	8 to 10	G	V
Nun lob, mein Seel, den Herren	136	17 to 19	Ab	V
Nun lob, mein Seel, den Herren	137	11 to 14	Ab	V
Nun lob, mein Seel, den Herren	137	33 to 36	Bb	V
Nun preiset alle Gottes Barmherzigkeit	138	15 to 18	B	V
Ich dank dir schon durch deinen Sohn	96	13 to 16	C	vi
Ich hab mein Sach Gott heimgestellt	98	7 to 9	C#	vi
Jesu, der du meine Seele	100	6 to 8	D	vi
Jesu, der du selbsten wohl	102	3 to 4	Eb	vi
Jesu, der du selbsten wohl	102	15 to 16	E	vi
Jesu, du mein liebstes Leben	103	5 to 8	F	vi
Jesu, Jesu, du bist mein	104	7 to 8	F#	vi
Jesu, meine Freude	105	7 to 8	G	vi
Jesu meiner Seelen Wonne	106	3 to 4	Ab	vi
Jesu meiner Seelen Wonne	106	11 to 12	Ab	vi
Jesu meiner Seelen Wonne	107	3 to 4	Bb	vi
Jesu meiner Seelen Wonne	107	11 to 12	B	vi
Das walt Gott Vater und Gott Sohn	37	6 to 8	C	vii°
Den Vater dort oben	39	13 to 14	C#	vii°
Allein Gott in der Höh sei Ehr	8	8 to 10	D	vii°
Alle Menschen müssen sterben	10	11 to 12	Eb	vii°
Ein feste Burg ist unser Gott	49	11 to 12	E	vii°
Ein feste Burg ist unser Gott	50	10 to 11	F	vii°
An Wasserflüssen Babylon	15	15 to 17	F#	vii°
Auf, auf, mein Herz	16	10 to 11	G	vii°
Es ist gewißlich an der Zeit	54	8 to 10	Ab	vii°
Es spricht der Unweisen Mund wohl	55	9 to 10	Ab	vii°
Da der Herr Christ zu Tische saß	32	11 to 12	Bb	vii°
Dank sei Gott in der Höhe	34	10 to 12	B	vii°

Appendix C. Musical Training Questionnaire

Participant ID: ………………….
Age: _______________
Sex: _______________
Handedness (left or right): _______________
Years of education: _______________
Do you consider yourself a musician? Yes/No
Do you **currently** play or are learning to play a musical instrument? Yes/No
If yes:

Instrument	Number of years played	Accumulated practice time	Start age

Have you ever, **in the past,** played or learned to play an instrument **but have stopped** playing it? Yes/No

If yes:

Instrument	Number of years played	Accumulated practice time	Start age	End age

On a scale of 1 to 5, rate the following
(1=None or Not Able; 2=Limited; 3=Average; 4=Above Average; 5=Extensive or Very Able).

Knowledge of music history:	1	2	3	4	5
Knowledge of music theory:	1	2	3	4	5
Ability to read music:	1	2	3	4	5
Overall music ability:	1	2	3	4	5

Other comments to music-related activities:
____
____

How many hours do you listen to music per week? _______________

Appendix D. Language Experience Questionnaire

Participant ID: ………………….

1. Are you bilingual in the sense that you are comfortable communicating with fluency in two or more languages? Yes/No

First language: _________________
Second Language: _________________
Other Languages: _________________

2. If you have answered YES to question 1, how proficient would you rate your ability in your languages? Please use a scale from 1 (not proficient) to 10 (native-like ability).

First language: _________________
Second Language: _________________
Other Languages: _________________

3. At what ages did you start learning your languages? (for your native language, please write 0).

First language: _________________
Second Language: _________________
Other Languages: _________________

5. How often have you switched between languages in the last year? (This includes speaking, writing, reading or listening to the different languages.)

On a daily basis On a weekly basis On a monthly basis

6. In the last year, what percentage of time have you used your different languages?

First language: _________________ (%)
Second Language: _________________ (%)
Other Languages: _________________ (%)

7. Have you experienced interference between languages in your day-to-day life?

(e.g., unintentionally saying a word in your first language while speaking your second language or feeling like you need to inhibit one of the languages while engaging with the other)

Yes, I have experienced language interference / No, I have not experienced language interference

Appendix E. EEG preprocessing details

Preprocessing was done in EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014). The protocol of the steps used is discussed below. The steps were:

1. High-pass filter (0.1Hz)
2. Epoch –200 +1000 ms (relative to the onset of the target word/chord)
3. Remove “bad” channels
4. Reference to Cz
5. Artefact correction using independent components analysis (ICA) and ADJUST
6. Artefact rejection using moving peak-to-peak window method in ERPLAB
7. Baseline correction (–200 to 0 ms)
8. Interpolation
9. Re-reference to average reference
10. Grand average
11. Low-pass filter (30Hz)

These steps are each discussed, in turn, below:

1. High-pass filtering

For EEG data, filtering choices involve compromising between attenuation of the signal and reducing noise. A 0.1 Hz filter was applied as filters at this frequency attenuate skin potentials and other slow drifts, and tend not to disrupt ERP effects(Kappenman &

Luck, 2012). In addition to reducing noise, high-pass filtering at 0.1 Hz aids artefact correction with ICA (step 5). High-pass filtering must be done before epoching to avoid edge artefacts in the epochs.

2. Epoch

Epochs were cut between –200 ms and +1000 ms relative to the target words/chords. Epoching at this stage aids subsequent preprocessing steps, as it makes file sizes smaller and more manageable. Moreover, it focuses preprocessing on time periods of interest, which makes bad electrode identification and artefact correction/detection more efficient as, when participants are not doing the task (e.g., before the experiment begins, or during breaks in between experimental blocks), they often speak or move in ways that affect the EEG signal that are not relevant when analysing task periods.

3. Removing bad electrodes

“Bad” electrodes occur if the quality of the connection between the electrode and the scalp is poor, or if there is something wrong with the electrode. The data that is collected from “bad” electrodes is usually not useful and leaving them in the data can hinder artefact correction (step 5) and artefact detection (step 6), it could also introduce noise into all channels during average re-referencing (step 9). “Bad” electrodes can be identified, as their activity is likely to be very different to the electrodes that are near them, or the log of their power spectral density ($\mu V^2/Hz$) can be much higher than all the other electrodes (suggesting that they are mostly picking up noise). EEGLAB has an automatic inbuilt electrode rejection function that uses a kurtosis method to identify potential “bad” electrodes. For the current experiments, this function was run, with the default standard deviation > 5 to be labelled as “bad”. The results were manually inspected, alongside channel activity and channel power spectra plots. Taking the results from all these methods into consideration, “bad” electrodes were then removed from the dataset.

4. Re-referencing

For the current ERP experiments, the data were first re-referenced to Cz. Re-referencing of Biosemi data (which was the form of the raw EEG data) is required at this stage by EEGLAB, and Cz was chosen because it is near the original reference channels of the Biosemi system. Later, in step 9, the average reference was applied. Average referencing is not advisable at this early stage as it can distort the ICA results (used for artefact correction). This is because the average reference does not result in a measure of absolute voltage at electrode sites. Instead, it is based on the pattern of differences in voltage amongst electrodes.

5. Artefact correction

The question of how to deal with artefacts is often discussed in EEG research (Luck, 2014). Some recommend that researchers ask participants to control their eye blinks (e.g., not blink when the target stimulus is presented). However, this can distract from the primary task, and participants might become more aware of their eye blinks and blink more than they would have. Ochoa and Polich (2000) found that the P3 in their study was smaller in amplitude and had a longer latency when participants were asked not to blink. Hence, for the current experiments, participants were not asked to control their eye blinks. Instead, they were told to relax and sit up straight in the chair, as this can reduce muscle artefacts from, e.g., facial movements, teeth grinding and neck strain.

For the current ERP experiments, artefact correction with ICA (independent component analysis) was applied before artefact rejection to correct for eye blinks, horizontal eye movements and muscle artefacts. It is also possible to not do artefact correction and to instead just do artefact rejection. However, this can result in much fewer trials as participants tend to blink during trials, which means that the data has decreased signal-to-noise ratio, making it difficult to detect ERP effects. Additionally, artefact rejection without artefact correction can result in a greater imbalance in the number of trials in each condition if, for example, participants happen to blink more in one condition than another. If this occurs, the imbalance between the number of trials in each condition could account for statistical differences between them. Therefore, for the current ERP analyses, artefact correction with ICA was applied before artefact rejection.

ICA is a suitable method for identifying eye blinks and horizontal eye movements, as these have a distinct scalp topography and other characteristics, such as sudden large voltage changes. It can also detect EMG noise due to movement of (or tension in) the jaw (this is usually detected in the T7 and T8 electrodes, which sit directly above the temporalis muscles that control the jaw). This type of EMG noise also has distinct characteristics and is straightforward to identify with ICA. A conservative approach to correcting ICA components was adopted. The automatic algorithm ADJUST was used to identify artefact components (Mognon et al., 2011). ADJUST identified eye blinks, eye movements and generic discontinuities in the data. It does this by combining stereotypical artefact-specific spatial and temporal features and aims to minimise disruption to neural signal. The components identified by ADJUST were visually inspected, and a decision was made about whether the component was artefactual or not. During the manual inspection, some artefactual components were identified that had not been identified by the ADJUST algorithm. These were also marked for correction. If there was doubt about whether a component included neural signal, it was not marked for correction.

6. Artefact rejection

The moving peak-to-peak window artefact rejection method from ERPLAB was used to reject additional trials that contained eye movements. Peak-to-peak amplitude is the difference between the most positive and most negative voltages within a window. A moving window peak-to-peak amplitude function computes the peak-to-peak amplitude within a series of windows within each epoch. The algorithm finds the largest peak-to-peak amplitude from within a specific time window (for the current experiments, this was set at 100 ms) in each epoch of data and then compares this largest value with a threshold value (which for the current experiment was kept at the default, 100 µV). Then, it moves 100 ms forwards and compares against the threshold again. It continues to do this until the end of the epoch. Trials are marked for rejection if the largest value exceeds the set threshold (100 µV). All trials marked for rejection with this method were reviewed manually before rejecting them, but the vast majority marked by this method was rejected.

7. Baseline correction

Baseline correction controls for slow drifts caused by factors like skin hydration, static electrical charges and skin potentials. The voltage during the pre-stimulus period usually provides a reasonable estimate of the voltage offset for that trial. It contains the offset but should not contain the stimulus-elicited ERP activity.

For the current ERP experiments, the default baseline correction was between – 200 to 0 ms relative to the target word or chord's onset. This baseline period of 200 ms allows for two full cycles of the alpha oscillation, which minimises artefactual effects of alpha activity (often due to skin potentials). Compared to shorter baseline periods, it gives a more accurate estimate of the voltage offset because little noise blips cancel out when averaging over more points in the baseline. Furthermore, Luck (2014) suggests that the pre-stimulus baseline period should be at least 20% of the overall epoch duration, and recommends a 200 ms pre-stimulus baseline with an 800 ms post-stimulus period.

Baseline correction should not be done before ICA for artefact correction, as it can affect the characteristics of independent components, making it less clear whether they represent artefacts or not (Groppe et al., 2009). The artefact rejection method chosen for the current experiments (moving peak-to-peak window) is not influenced by baseline correction. Thus, although baseline correction was done after artefact rejection, it could have been done immediately before artefact rejection and given the same results.

8. Interpolation

“Bad” channels that were removed in step 3 were interpolated. It is important to do this before re-referencing to the average reference so that the grand average is calculated based on the same amount of information for all participants.

9. Re-reference to average reference

After interpolation, the data were re-referenced to the average reference. For all participants, the 64 scalp channels contributed to the average reference.

10. Grand averaging

At this stage, grand averages were taken for each condition. Luck (2014) suggests that ERP studies should be designed with between 30–40 trials per condition with 20 participants to have an acceptable signal-to-noise ratio for ERP component analysis. He gives these guidelines, assuming that between 10–25% of these will be rejected due to artefacts. The average number of trials included in the final analysis for each condition is presented in the Methods sections of each ERP experiment, under the Analysis headings. There were initially 48 trials per condition for all the current ERP experiments, and always more than 30 per condition on average in the final analysis. Therefore, all were assumed to have an acceptable signal-to-noise ratio.

11. Low-pass filtering (30 Hz)

Finally, a 30Hz low-pass filter was applied to attenuate line noise and EMG noise. It is crucial to apply low-pass filters after artefact rejection because they make it more difficult to identify artefacts, which are often characterised by high frequencies. It is advised to apply low-pass filters at this stage by Luck (2014), to allow for flexibility in future analyses.

Appendix F. Frequency of accumulated practice time/measuring musical expertise

Most previous research investigating expertise effects associated with musical training in adults has sampled highly trained, professional musicians. As a result, it is unclear whether lower levels of training are enough to produce reliable expertise effects. The current thesis aimed to investigate whether there are expertise effects in groups of amateur musicians compared to non-musicians. For Experiments 1 and 4, the recruitment adverts asked for "musicians" and "non-musicians". It was deemed crucial that there were no significant subjective differences in how the expectancy violations were rated as "Expected" and "Unexpected" based on musical training experience, to enable meaningful interpretations of any differences in ERP effects between groups. For Experiments 1 and 4, the participants' responses to the question "Do you consider yourself a musician? Yes/No" on the Musical Training Questionnaire (see Appendix C) was used to distinguish between groups[10].

For all other experiments, participants were not recruited based on musical training experience. This sampling method was chosen because it was thought that heightening participants' awareness of their musical ability before they completed the ERP tasks could have had additional effects that may have affected the ERP effects. As a result of this sampling method, participants tended to be less certain about whether they were a "musician" or a "non-musician", and so a more objective measure of their musical training experience was needed. Previous studies have used accumulated practice time to distinguish between musicians and non-musicians (Goldman et al., 2018; Jentzsch et al., 2014), and this was deemed a suitable method for the current experiments.

Next, came the challenge of choosing a suitable cut-off point in accumulated practice time to distinguish between amateur musicians and non-musicians. This was based on the median accumulated practice time of all participants after all the data was collected. Across the 132 participants who took part in the experiments, the median

---

[10] These matched up with the later–chosen criteria for distinguishing between amateur musicians and non-musicians, based on accumulated practice time.

accumulated practice time was 1055 hours. The raw data showed that one participant had 1040 hours of accumulated practice time, and the next lowest accumulated practice time was 855 hours. Therefore, 1000 hours was deemed most suitable as a cut-off between amateur musicians and non-musicians. Additionally, the cut-off at 1000 hours matched with the musician and non-musician grouping in Experiments 1 and 4. Across all experiments, the range of accumulated practice time was 0–885 hours for non-musicians (*N* = 65, *M* = 232 hours, *SD* = 885 hours) and 1040–13,728 hours (*N* = 67, *M* = 3597 hours, *SD* = 2588 hours) for amateur musicians. The frequency of accumulated practice time for all participants (data collapsed across the seven experiments) is shown in Figure F1 and is shown separately for each experiment in Figure F2.

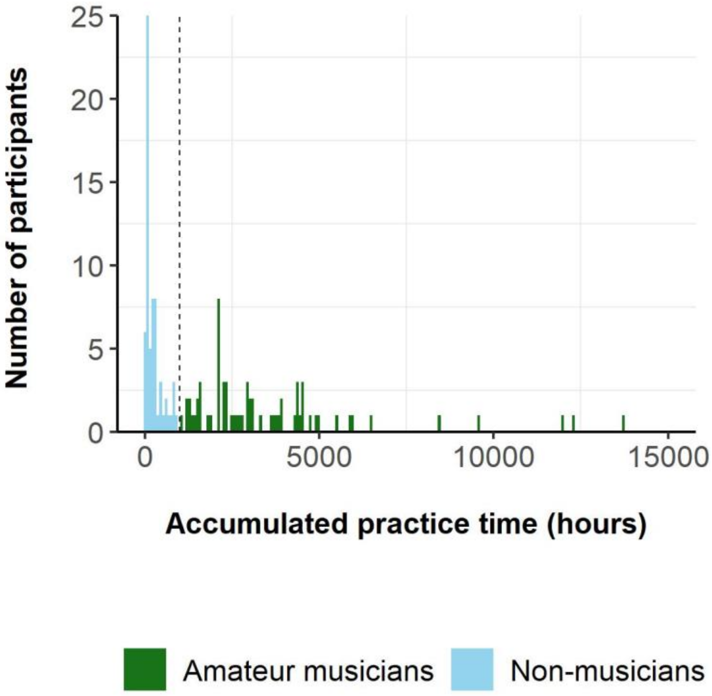


Figure F1 – Frequency of accumulated practice times for amateur musicians (green) and non-musicians (blue). The cut-off between amateur musicians and non-musicians was 1000 hours of accumulated practice time, marked by the dashed line.

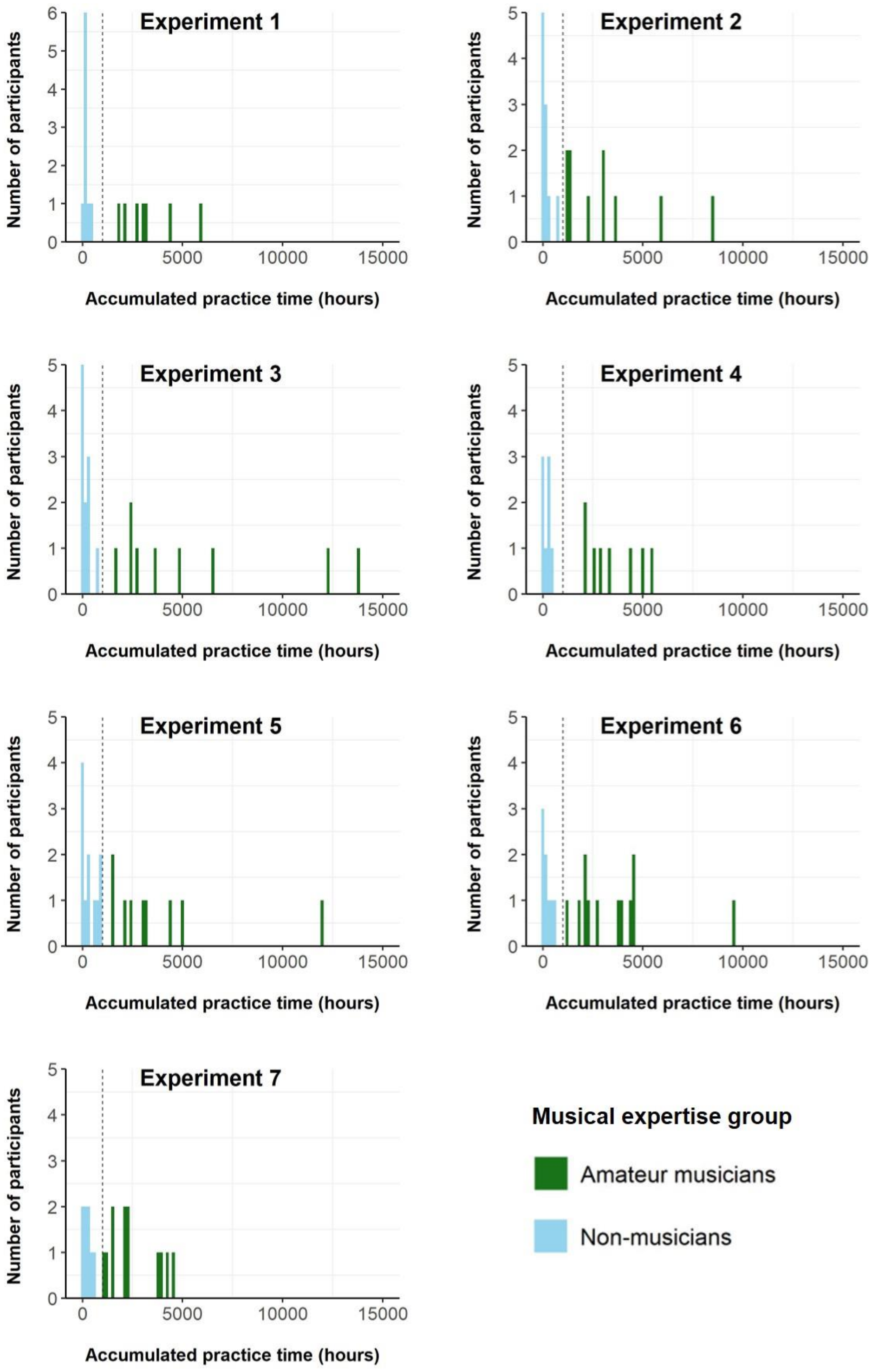


Figure F2 – Graphs showing the accumulated practice times by the frequency of participants in each experiment. The cut-off criteria at 1000 hours (> 1000 hours for amateur musicians and < 1000 hours for non-musicians) is marked with a dashed line.

Appendix G. Experiment 2: Results of ANOVA tests for distinct ERP comparisons

Table G1 – Results of ANOVA tests for the electrodes and time windows for all three manipulations, Experiment 2. The results for the N400 for semantic violations, P600 for grammar violations, and P300 for harmonic violations are shaded in grey.

Analysis	Semantic	Grammar	Harmony
Pz 300–450 ms	$F(2, 38) = 6.12$, $p = .005$, $\eta p^2 = 0.24$	$F(1,19) = 0.77$, $p = .390$, $\eta p^2 = 0.04$	$F(2, 38) = 2.49$, $p = .097$, $\eta p^2 = 0.12$
Cz 500–800 ms	$F(2, 38) = 2.64$, $p = .769$, $\eta p^2 = 0.14$	$F(1,19) = 17.82$, $p = .001$, $\eta p^2 = 0.48$	$F(2, 38) = 2.28$, $p = .116$, $\eta p^2 = 0.11$
Fz 250–400 ms	$F(2, 38) = 2.44$, $p = .101$, $\eta p^2 = 0.11$	$F(1,19) = 0.03$, $p = .859$, $\eta p^2 = 0.002$	$F(2, 38) = 10.19$, $p < .001$, $\eta p^2 = 0.35$

Appendix H. Experiment 3: Results of ANOVA tests for distinct ERP comparisons

Table H1 – Results of ANOVA tests for the electrodes and time windows for all three manipulations, Experiment 3. The results for the N400 for semantic violations, P600 for grammar violations, and P300 for harmonic violations. are shaded in grey.

Analysis	Semantic	Grammar	Harmony
Pz 300–450 ms	$F(2, 38) = 4.91$, $p = .013$, $\eta p^2 = 0.21$	$F(1,19) = 0.10$, $p = .757$, $\eta p^2 = 0.005$	$F(2, 38) = 1.90$, $p = .163$, $\eta p^2 = 0.09$
Cz 500–800 ms	$F(2, 38) = 0.37$, $p = .692$, $\eta p^2 = 0.19$	$F(1, 19) = 4.82$, $p = .041$, $\eta p^2 = 0.20$	$F(2, 38) = 2.82$, $p = .086$, $\eta p^2 = 0.12$
Fz 250–400 ms	$F(2, 38) = 0.96$, $p = .391$, $\eta p^2 = 0.05$	$F(1,19) = 0.44$, $p = .517$, $\eta p^2 = 0.02$	$F(2, 38) = 5.41$, $p = .009$, $\eta p^2 = 0.22$

Appendix I. Baseline comparisons: Experiment 5

Language task: Interstimulus interval effect

Table I1 – Comparison of the N1 and P2 interstimulus interval effects between the two baseline analyses: –200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey), in the language task.

					Mean amplitude µV (*SD*)		
Test	Baseline (ms)	F test	Pairwise comparison	p	Short 500 ms	Regular 1000 ms	Long 1500 ms
N1: Fz, 80–140 ms	–200 to 0	*F*(2, 38) = 3.76, *p* = .033, ηp^2 = 0.17	Regular and short	.009	–0.87 (0.17)	–014 (0.19)	–0.74 (0.32)
			Regular and long	.316			
			Long and short	1.00			
	0 to +50	*F*(2, 38) = 1.21, *p* = .310, ηp^2 = 0.06	Regular and short	.384	–0.61 (0.23)	–0.28 (0.21)	–0.54 (0.22)
			Regular and long	.768			
			Long and short	1.00			
P2: Pz, 140–260 ms	–200 to 0	*F*(2, 38) = 10.01, *p* < .001, ηp^2 = 0.04.	Regular and short	.006	1.37 (0.40)	0.26 (0.25)	1.56 (0.32)
			Regular and long	.001			
			Long and short	1.00			
	0 to +50	*F*(2, 38) = 5.65, *p* = .007, ηp^2 = 0.23	Regular and short	.021	1.25 (0.37)	0.48, (0.28)	1.18 (0.26)
			Regular and long	.011			
			Long and short	1.00			

Music task: Interstimulus interval effect

Table I2 – Comparison of the N1 and P2 interstimulus interval effects between the two baseline analyses: –200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey), in the music task.

					Mean amplitude µV (*SD*)		
Test	Baseline (ms)	*F* test	Pairwise comparison	p	Short 500 ms	Regular 1000 ms	Long 1500 ms
N1: Fz, 80–140 ms	–200 to 0	*F*(2, 38) = 7.45, *p* = .002, ηp^2 = 0.28	Regular and short	.018	–0.42 (0.30)	0.35 (0.18)	–0.57 (0.23)
			Regular and long	.009			
			Long and short	1.00			
	0 to +50	*F*(2, 38) = 6.63, *p* = .003, ηp^2 = 0.26	Regular and short	.011	–0.50 (0.21)	0.16 (0.16)	–0.50 (0.25)
			Regular and long	.027			
			Long and short	1.00			
P2: Pz, 140–260 ms	–200 to 0	*F*(2, 38) = 8.47, *p* = .001, ηp^2 = 0.31.	Regular and short	.002	1.20 (0.25)	0.12 (0.85)	0.85 (0.26)
			Regular and long	.052			
			Long and short	.641			
	0 to +50	*F*(2, 38) = 6.78, *p* = .003, ηp^2 = 0.26	Regular and short	.012	1.05 (0.24)	0.11 (0.13)	0.66 (0.25)
			Regular and long	.125			
			Long and short	.304			

Data collapsed across language and music tasks: Interstimulus interval effect

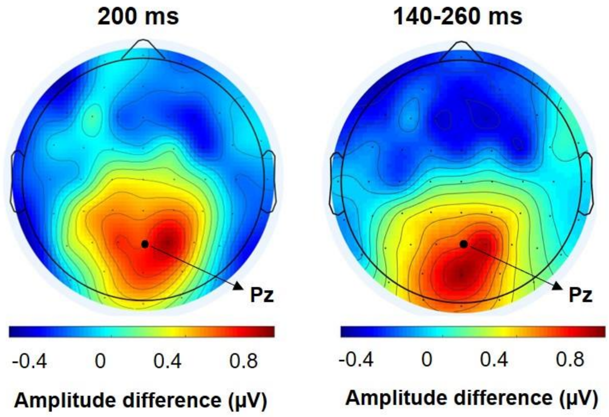
Table I3 – The P2 effects, with –200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey), when collapsed across the language task and the music task.

					Mean amplitude µV (*SD*)		
Test	Baseline (ms)	*F* test	Pairwise comparison	*p*	Short 500 ms	Regular 1000 ms	Long 1500 ms
P2: Pz, 140–260 ms	–200 to 0	*F*(2, 38) = 16.83, *p* < .001, ηp^2 = 0.47	Regular and short	< .001	1.30 (0.23)	0.20 (0.16)	1.22 (0.14)
			Regular and long	< .001			
			Short and long	1.00			
	0 to +50	*F*(2, 38) = 10.68, *p* < .001, ηp^2 = 0.36	Regular and short	.001	1.15 (0.20)	0.30 (0.15)	0.92 (0.13)
			Regular and long	.003			
			Short and long	.859			

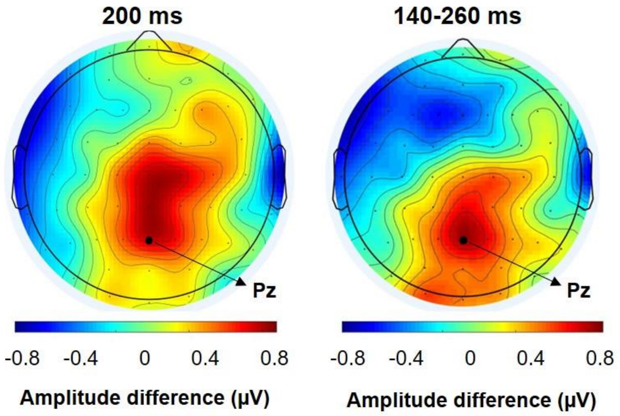
Appendix J. P2 scalp topography maps – method comparison

A visual scalp topography map comparison of the P2 effect between the instantaneous amplitude method at 200 ms (which is included in the experimental write-ups, shown on the left below) and the mean amplitude method (between 140–260 ms, shown on the right). For each P2 effect presented in this thesis, both methods are presented in this Appendix, for completeness. The difference was calculated as the irregular (the average of the short 500 ms and long 1500 ms) interstimulus intervals minus the regular 1000 ms interstimulus intervals. As in the main text, the -200 to 0 ms baselines are plotted.

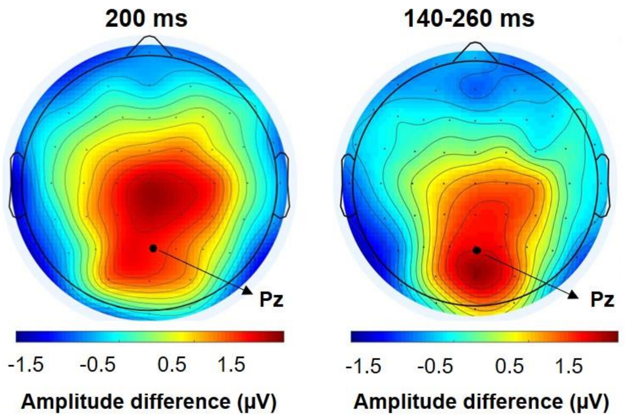
Experiment 5: Language task



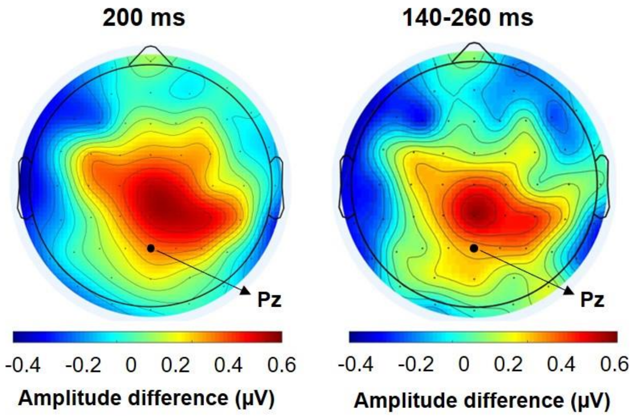
Experiment 5: Music task



Experiment 6 (language task)



Experiment 7 (music task)



Appendix K. Baseline comparisons: Experiment 6 (Language task)

Interstimulus interval effects

Table K1 – Results of the P2 interstimulus interval effect, –200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey), when collapsed across i) standard word and unrelated semantic violations and ii) standard word and grammar expectancy violations.

				Mean amplitude µV (*SD*)	
Test	Analysis	Baseline (ms)	*F* test	Regular (1000 ms)	Irregular (500/1500 ms)
P2: Pz, 140–260 ms	Semantic	–200 to 0	$F(1, 19) = 14.03$, $p = .001$, $\eta p^2 = 0.43$	0.53 (0.16)	1.40 (0.24)
		0 to +50	$F(1, 19) = 2.15$, $p = .159$, $\eta p^2 = 0.10$	0.89 (0.16)	1.16 (0.19)
	Grammar	–200 to 0	$F(1, 19) = 18.10$, $p < .001$, $\eta p^2 = 0.49$	0.55 (0.18)	1.50 (0.26)
		0 to +50	$F(1, 19) = 5.57$, $p = .029$, $\eta p^2 = 0.23$	0.87 (0.17)	1.24 (0.17)

Semantic and grammar effects

Table K2 – Sentence end type main effect results for the two baseline analyses (–200 to 0 ms, no shading, and 0 to +50 ms, shaded in grey). The data is collapsed across interstimulus interval types (regular/irregular).

				Mean amplitude µV (*SD*)	
Test	Analysis	Baseline (ms)	*F* test	Standard	Word violation
N400: Pz, 300–450 ms	Semantic	–200 to 0	$F(1, 19) = 5.92$, $p = .025$, $\eta p^2 = 0.24$	0.27 (0.27)	–0.23 (0.26)
		0 to +50	$F(1, 19) = 7.47$, $p = .013$, $\eta p^2 = 0.28$	0.39 (0.22)	–0.22 (0.29)
P600: Cz, 500–800 ms	Grammar	–200 to 0	$F(1, 19) = 4.63$, $p = .045$, $\eta p^2 = 0.20$	–0.73 (0.30)	–0.04 (0.17)
		0 to +50	$F(1, 19) = 8.84$, $p = .008$, $\eta p^2 = 0.32$	–0.68 (0.26)	0.002 (0.14)

Appendix L. Interaction results: Experiment 6 (Language task)

Interaction results

Table L1 – Experiment 6 Interactions between sentence end type and interstimulus interval, for –200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey) analyses.

Test	Analysis	Baseline (ms)	*F* test
P2: Pz, 140–260 ms	Semantic	–200 to 0	F(1, 19) = 0.96, *p* = .341, ηp2 = 0.05
		0 to +50	F(1, 19) = 0.25, *p* = .625, ηp2 = 0.013
	Grammar	–200 to 0	F(1, 19) = 0.27, *p* = .607, ηp2 = 0.01
		0 to +50	F(1, 19) = 0.08, *p* = .776, ηp2 = 0.004
N400: Pz, 300–450 ms	Semantic	–200 to 0	F(1, 19) = 0.08, *p* = .785, ηp2 = 0.004
		0 to +50	F(1, 19) = 0.001, *p* = .927, ηp2 < 0.01
P600: Cz, 500–800 ms	Grammar	–200 to 0	F(1, 19) = 0.32, *p* = .580, ηp2 = 0.02.
		0 to +50	F(1, 19) = 0.002, *p* = .969, ηp2 < 0.01

Appendix M. Baseline comparisons: Experiment 7 (Music task)

Interstimulus interval effects

Table M1 – Interstimulus interval main effects for the two baseline analyses, –200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey). The data is averaged across harmonic chord progression end types.

			Mean amplitude µV (*SD*)	
Test	Baseline (ms)	*F* test	Regular (1000 ms)	Irregular (500/1500 ms)
N1: Fz, 80–140 ms	–200 to 0	*F*(1, 19) = 5.03, *p* = .037, *ηp*^{2} = 0.21	–0.01 (0.18)	–0.51 (0.17)
	0 to +50	*F*(1, 19) = 7.02, *p* = .016, *ηp*^{2} = 0.27	–0.09 (0.13)	–0.48 (0.13)
P2: Pz, 140–260 ms	–200 to 0	*F*(1, 19) = 10.34, *p* = .005, *ηp*^{2} = 0.35	0.27 (0.17)	0.74 (0.20)
	0 to +50	*F*(1, 19) = 3.29, *p* = .086, *ηp*^{2} = 0.15	0.39 (0.15)	0.70 (0.20)

Harmony effects

Table M2 – Harmonic expectancy violation main effects for the two baseline analyses, for –200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey). The data is averaged across interstimulus interval types.

			Mean amplitude µV (*SD*)	
Test	Baseline (ms)	*F* test	Tonic (I)	Unrelated harmony
P300: Fz, 250–400 ms	–200 to 0	*F*(1, 19) = 4.10, *p* = .057, *ηp*^{2} = 0.35	–0.52 (0.20)	–0.05 (0.24)
	0 to +50	*F*(1, 19) = 3.22, *p* = .089, *ηp*^{2} = 0.15	–0.49 (0.18)	–0.14 (0.20)

Appendix N. Interaction results: Experiment 7 (Music task)

Interaction results

Table N1 – Experiment 7 Interactions between chord progression end type and interstimulus interval, for –200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey) baseline analyses.

Test	Baseline (ms)	*F* test
N1: Fz, 80–140 ms	–200 to 0	$F(1, 19) = 0.62$ $p = .443$, $\eta p^2 = 0.03$
	0 to +50	F(1, 19) = 0.39, *p* = .541, ηp2 = 0.20
P2: Pz, 140–260 ms	–200 to 0	$F(1, 19) = 0.31$, $p = .587$, $\eta p^2 = 0.02$
	0 to +50	F(1, 19) = 1.08, *p* = .311, ηp2 = 0.05
P300: Fz, 250–400 ms	–200 to 0	$F(1, 19) = 0.96$, $p = .340$, $\eta p^2 = 0.05$
	0 to +50	F(1, 19) = 0.004, *p* = .949, ηp2 < 0.01

Appendix O. P2 results comparison between Experiments 6 and 7

Interstimulus interval effect

Table O1 – Interstimulus interval main effects for the two baseline analyses, –200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey). The data is averaged across sentence end types and harmonic end types to aid comparisons between language and music tasks.

			Mean amplitude µV (*SD*)	
Test	Baseline (ms)	*F* test	Regular (1000 ms)	Irregular (500/1500 ms)
P2: Pz, 140–260 ms	–200 to 0	*F*(1, 38) = 27.27, *p* < .001, ηp^2 = 0.42	0.43 (0.12)	1.10 (0.15)
	0 to +50	*F*(1, 38) = 5.36, *p* = .032, ηp^2 = 0.22	0.59 (0.09)	0.95 (0.12)

Interaction result

Table O2 – Interaction results between interstimulus interval (regular/irregular) and task (language/music) for the two baseline analyses-200 to 0 ms (no shading) and 0 to +50 ms (shaded in grey).

Test	Baseline (ms)	*F* test
P2 interaction, interstimulus interval * task. Pz, 140–260 ms	–200 to 0	*F*(1, 38) = 2.40, *p* = .130, ηp^2 = 0.06
	0 to +50	*F*(1, 38) = 0.14, *p* = .711, ηp^2 < .001

Appendix P. Ethical approval forms

Ethical approval for Experiment 1

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

26 February 2018

Dear Joanna

Thank you for submitting your ethical application which was considered by the School of Psychology & Neuroscience Ethics Committee on 22nd February 2018; the following documents have been reviewed:

1. Ethical Application Form
2. Advertisement
3. Participant Information Sheet
4. Participant Consent Form: Anonymous Data
5. Participant Debriefing Form
6. Music Experience Questionnaire
7. Data Management Plan

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	PS13356	**Approved on:**	26/02/2018	**Approval Expiry:**	26/02/2023
Project Title:	Do non-musicians detect atypical musical chord progression?				
Researcher:	Joanna Moodie				
Supervisor:	Dr Ines Jentzsch				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Shona M Duignan

Convener of the School Ethics Committee

cc Dr Ines Jentzsch (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

The University of St Andrews is a charity registered in Scotland: No SC013532

Ethical approval for Experiments 2 and 3

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

10 January 2018

Dear Joanna

Thank you for submitting your ethical application which was considered at the School of Psychology & Neuroscience Ethics Committee meeting on 14th December 2017; the following documents have been reviewed:

1. Ethical Application Form
2. Advertisement
3. Participant Information Sheet
4. Participant Consent Form: Anonymous Data
5. Participant Debriefing Form
6. Questionnaires ('Music Experience' and 'Language Experience and Proficiency')
7. Data Management Plan

The School of Psychology & Neuroscience Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	PS13256	**Approved on:**	03/01/2018	**Approval Expiry:**	03/01/2023
Project Title:	Musical and linguistic expectancy violations				
Researcher:	Joanna Moodie				
Supervisor:	Dr Ines Jentzsch				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee

cc Dr Ines Jentzsch (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

Ethical approval for Experiments 4 and 5

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

14 February 2019

Dear Joanna

Thank you for submitting your amendment application which comprised the following documents:

1. Ethical Amendment Application Form
2. Advertisements: Pilot Study and EEG Study
3. Participant Information Sheets: Pilot Study and EEG Study
4. Participant Consent Form
5. Participant Debriefing Form
6. Questionnaires: Music Experience and Language Background

The School of Psychology & Neuroscience Ethics Committee is delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has approved this ethical amendment application. The particulars of this approval are as follows –

Original Approval Code:	PS13256	**Approved on:**	03/01/2018
Amendment Approval Date:	07/02/2019	**Approval Expiry Date:**	03/01/2023
Project Title:	Music and language-based expectancy		
Researcher:	Joanna Moodie		
Supervisor:	Dr Ines Jentzsch		

Ethical amendment approval does not extend the originally granted approval period of five years, rather it validates the changes you have made to the originally approved ethical application. If you are unable to complete your research within the original five-year validation period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the understanding that you adhere to the 'Guidelines for Ethical Research Practice' (http://www.st-andrews.ac.uk/media/UTRECguidelines%20Feb%2008.pdf).

Yours sincerely

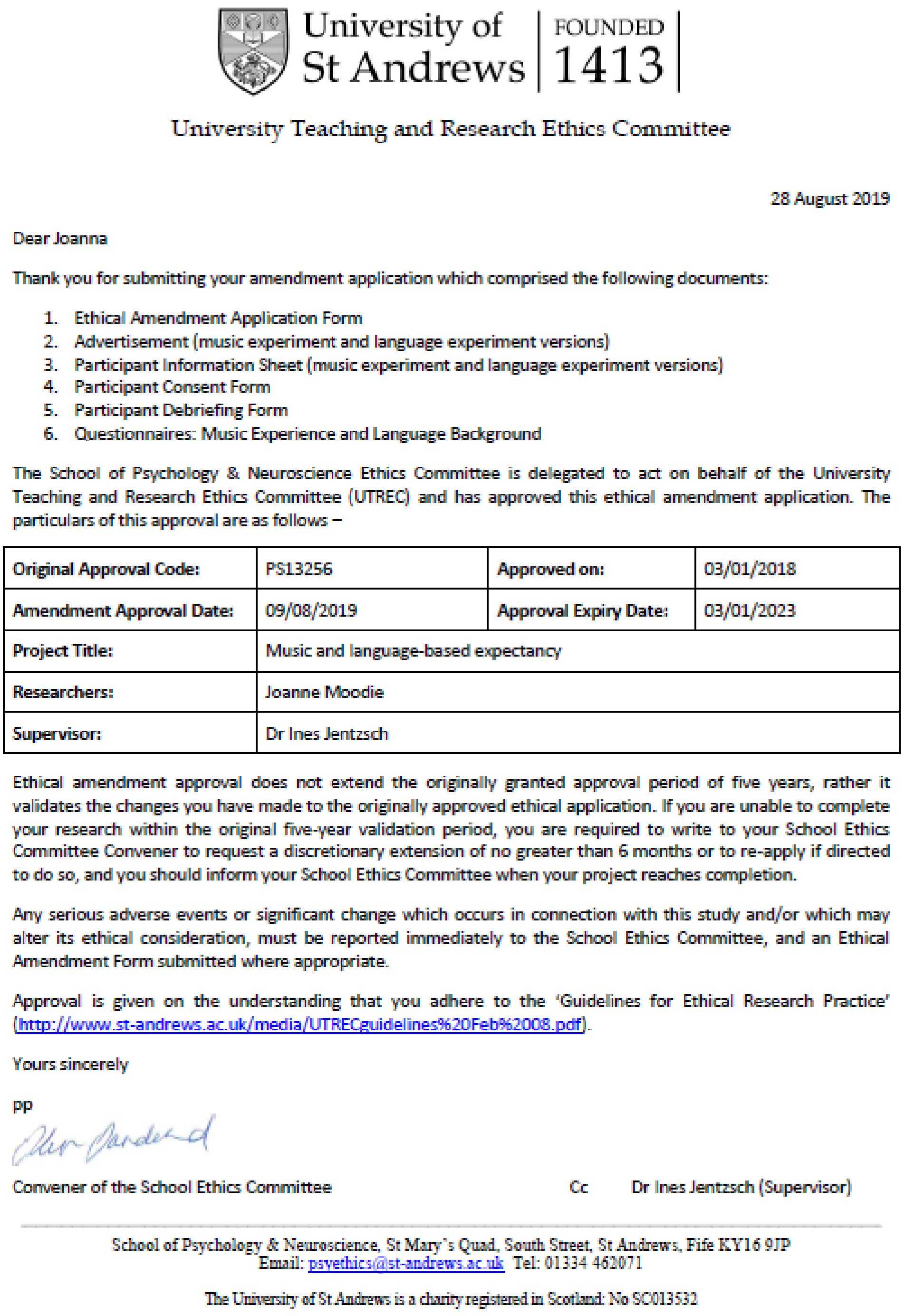
Convener of the School Ethics Committee

cc Dr Ines Jentzsch (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

The University of St Andrews is a charity registered in Scotland: No SC013532

Ethical approval for Experiments 6 and 7



University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

28 August 2019

Dear Joanna

Thank you for submitting your amendment application which comprised the following documents:

1. Ethical Amendment Application Form
2. Advertisement (music experiment and language experiment versions)
3. Participant Information Sheet (music experiment and language experiment versions)
4. Participant Consent Form
5. Participant Debriefing Form
6. Questionnaires: Music Experience and Language Background

The School of Psychology & Neuroscience Ethics Committee is delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has approved this ethical amendment application. The particulars of this approval are as follows –

Original Approval Code:	PS13256	**Approved on:**	03/01/2018
Amendment Approval Date:	09/08/2019	**Approval Expiry Date:**	03/01/2023
Project Title:	Music and language-based expectancy		
Researchers:	Joanne Moodie		
Supervisor:	Dr Ines Jentzsch		

Ethical amendment approval does not extend the originally granted approval period of five years, rather it validates the changes you have made to the originally approved ethical application. If you are unable to complete your research within the original five-year validation period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

Any serious adverse events or significant change which occurs in connection with this study and/or which may alter its ethical consideration, must be reported immediately to the School Ethics Committee, and an Ethical Amendment Form submitted where appropriate.

Approval is given on the understanding that you adhere to the 'Guidelines for Ethical Research Practice' (http://www.st-andrews.ac.uk/media/UTRECguidelines%20Feb%2008.pdf).

Yours sincerely

pp

Convener of the School Ethics Committee

Cc Dr Ines Jentzsch (Supervisor)

School of Psychology & Neuroscience, St Mary's Quad, South Street, St Andrews, Fife KY16 9JP
Email: psyethics@st-andrews.ac.uk Tel: 01334 462071

The University of St Andrews is a charity registered in Scotland: No SC013532