Statistical learning and auditory processing in adults and children with music training: a behavioural and ERP study

by

Pragati Rao Mandikal Vasuki

BSc. (Speech and Hearing), MSc. (Audiology)

A thesis submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Linguistics (Audiology section),

The HEARing CRC &

ARC Centre of Excellence in Cognition and its Disorders

Macquarie University, Sydney, Australia

September, 2016

MACQUARIE University SYDNEY · AUSTRALIA

Declaration

I certify that the research presented in the thesis is my original work conducted as part of my PhD research under the supervision of A/Prof Mridula Sharma, A/Prof Joanne Arciuli, and Distinguished Prof Katherine Demuth. All sources have been acknowledged and my contribution is clearly identified in the thesis. This work has not been submitted for a higher degree to any other university or institution. Ethics approval for this PhD project was obtained from the Human Research Ethics Committee, Macquarie University (Reference No. 5201300459).

Chapter 2 from this thesis has been published as Mandikal Vasuki, P.R., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research*, 342, 112-123

Pragati Rao Mandikal Vasuki

This page is intentionally left blank

Acknowledgements

The past three and a half years have been like a roller coaster ride with its many ups and downs. The first big high was to receive the Australia development scholarships and the prestigious Australia Leadership Award, which made my dream of an international educational experience come true. The scholarship and its perks have definitely helped in making this PhD journey memorable. I am also grateful for the support of the HEARing CRC and ARC Centre of Excellence of Cognition and its disorders during my candidature.

No words can express my gratitude towards my primary supervisor Mridula Sharma. Mridula, you have been an exceptional mentor with your boundless energy and limitless patience. My heartfelt gratitude for helping me achieve a full-rounded PhD experience. You were always willing to offer advice and counsel for any problems that I encountered (academic & non-academic). It has been a real privilege working with you.

My sincere gratitude to my adjunct supervisor Joanne Arciuli who introduced me to the fascinating world of statistical learning. I have the highest appreciation for your support and generosity in sharing knowledge. Thanks are also due to my associate supervisor Katherine Demuth for her encouragement and feedback on my work. In particular, the diversity and balance of skills of my supervisors in areas such as audiology, electrophysiology, psycholinguisitcs, and cognition have enabled me to grow my knowledge in several topics.

My special thanks to all the participants and parents for bringing in the children for the lengthy testing sessions.

Thanks to Ronny Ibrahim for the wonderful discussions we've shared. It has been a pleasure working with you and learning the nuances of advanced EEG processing with you.

A special thanks to Rohith for helping me during this arduous journey. Your constructive criticism and logical thinking have helped hone my skills as a scientist. I am grateful for all the advice and words of encouragement.

My heartfelt gratitude to Vishnu Nair for being an amazing friend and mentor during his time at Macquarie. Vishnu, you assisted in my journey to Sydney as well as offered your friendship and care when I most needed it.

I would like to thank my wonderful colleagues at Macquarie University for their friendship and support. I am grateful to the staff at the Audiology section for the delightful discussions and constructive feedback on my work. In particular, I am grateful to the many members of Audiology and Hearing Research group as well as the Child Language Lab for providing me with valuable advice for my presentations. I thank (in alphabetical order) Adam, Amanda, Baljeet, Bram, Carmen, Chevelle, Chi, Fabrice, Fadwa, Heivet, Isabelle, Ivan, Jaime, Joaquin, John, Kelly, Mariana, Matthieu, Nan, Nick, Rakshita, Shivali and Yatin. A special cheers to Chris for being a wonderful friend and officemate.

Thanks to my wonderful friends. Our shared laughter and trips made my stay in Sydney very memorable. Thanks Amol, Imam, Ishan, Karthik, Ketan, Manju, Murali, Nandini, Nitin, Rashid and Shalini.

I would like to thank Vishal and Ben Go for encouraging me to keep going during these last few months.

My family has been my pillar of support during this PhD adventure. I am grateful to have parents who have always encouraged me to reach out for the stars. Life in Sydney would not have been bearable if I did not have the wonderful support of my sister, jeejaji and little Anushaa.

I wish to thank all the people who have believed in me, assisted me, and encouraged me to keep trying. Without the special people in my life who shared my smiles and sorrows, my life would have been poorer and this work would never be possible.

Table of contents

This page is intentionally left blank

Thesis abstract

Background: Statistical learning (SL) ability helps individuals to extract probabilistic regularities from input without conscious awareness. There is debate in the literature about whether musical expertise is linked with performance on SL tasks. This thesis investigated the association between musical training and a variety of auditory and cognitive tasks including SL.

Methods: Data from 40 adults (17 musicians) and 50 children (25 musically trained) were collected. Over the course of 4 experiments, auditory and cognitive measurements, as well as behavioural and online electrophysiological assessments of both auditory and visual SL were obtained in this population. SL was evaluated using the embedded triplet paradigm. Auditory processing measures such as frequency discrimination, dichotic listening tasks and cognitive measures such as the digit span task, sustained attention were also evaluated. In children, a measure of musical abilities was also obtained.

Results: Experiments showed significantly better auditory SL in individuals with music training. However, individuals with music training did not outperform the control group in visual SL task. Similar results were observed even in children with musical training who have had at least 1.5 years of music training after controlling for socio-economic status, parents' education background across the groups. Musically trained adults and children outperformed their untrained counterparts as they showed distinct responses (larger responses for initial stimulus of the triplets) in the online auditory SL task. Importantly, experiments showed that performance on SL tasks was independent of auditory and cognitive processing measures. Additionally, individual differences in musical abilities were related to the capacity for SL in children. The use of electrophysiological indices such as N1 and N400 as online measures of SL in the two modalities is discussed.

Conclusions: The findings add to the growing literature on the nature of association of music training and other skills such as SL, auditory and cognitive processing skills.

This page is intentionally left blank

Chapter 1

Introduction

"I would teach children music, physics and philosophy; but more importantly music; for in the patterns of music and all the arts, are the keys to learning" – Plato

Our world is filled with patterns and regularities. Each day, our senses are bombarded with an assortment of stimuli. We learn to categorize and classify these stimuli in order to understand the regularities in our environment. Understanding the regularities in our environment is one of the basic aspects of the human cognitive processing. Our brain detects regularities and uses these to make predictions and choose the appropriate behaviour. One of the mechanisms responsible for detection and utilization of regularities proceeds implicitly and is known as statistical learning (SL). The term SL was proposed by Saffran, Aslin, and Newport (1996) to describe infants' ability to identify word boundaries solely from the statistical relationships between syllables in a continuous stream of pseudo-speech. A growing body of evidence suggests that SL plays a key role in a wide variety of cognitive processes such as language acquisition (Kuhl, 2004), language comprehension (Misyak & Christiansen, 2007; Misyak, Christiansen, & Tomblin, 2010), perception of speech in adverse listening conditions (Conway, Bauernschmidt, Huang, & Pisoni, 2010), and music appreciation (Loui, 2012; Rohrmeier & Rebuschat, 2012). Indeed, the field of SL has expanded exponentially (Perruchet & Pacton, 2006) and is a topic gathering increasing interest in cognitive science research.

The field of SL has its roots in trying to understand how infants acquire language in a short period of time. It is well known that language has many statistical regularities. It has been suggested that SL helps in learning these regularities by facilitating various processes such as word segmentation, vocabulary learning, and learning of syntax (Gomez, 2002; Kidd & Arciuli, 2016; Saffran, Newport, & Aslin, 1996; Saffran & Wilson, 2003). Like language, music is also a complex and highly structured form of human communication which involves a range of cognitive processes like perception, parsing, and production (Rohrmeier & Rebuschat, 2012). Music, similar to language has many statistical regularities notably regularities concerning the frequencies of co-occurrence between musical events (i.e., notes, chords) (Krumhansl, 2001; Tillmann, Bharucha, & Bigand, 2000). It has been suggested that the learning mechanisms (such as SL) involved in the acquisition of language also underlie the acquisition of music (Loui, 2012). For example, listeners become sensitive to statistical regularities in music through mere exposure to music (Tillmann & McAdams, 2004). It is plausible that musical training facilitates the process of detection of statistical regularities. This introductory chapter gives an overview of the principles and mechanisms involved in SL. The chapter then examines the structural and functional consequences of music training with an emphasis on the existing literature connecting musical expertise and SL. The chapter concludes with an orientation to the structure of the remainder of this thesis.

1.1. What are the ‘statistics’ in statistical learning?

The study of SL has often been limited to the study of conditional relationships. This is because when the term SL was proposed, it was introduced as one of the mechanisms to understand how infants are able to extract words from continuous speech or the process of word segmentation/speech segmentation. To understand the relationship between musical practice and SL, the studies in this thesis investigate the learner’s sensitivity to conditional statistics. However, other research studies have shown that humans are sensitive to many different kinds of statistics. These statistics can be broadly divided into three categories: conditional, distributional, and, cue-based statistics (Thiessen, Kronstein, & Hufnagle, 2013).

The three categories are further discussed below, with more attention paid to conditional statistics owing to its direct bearing on the studies presented in this thesis.

1.1.1. Conditional statistics

Conditional statistics describe the predictive relationship between two events. For example, transitional probability, one of the most commonly used conditional statistics, describes the probability of occurrence of the event Y given the occurrence of event X. A frequentist[1] interpretation of probabilities would lead to an estimate shown in Equation 1.

$$p(Y|X) = \frac{\text{Number of occurences of Y followed by X}}{\text{Number of occurrences of X}}$$

Equation 1. Frequentist interpretation of transitional probability of X and Y

If Y regularly follows X, then the transitional probability is high; if Y rarely follows X, then it is low. Sensitivity to conditional relations is one of the basic processes involved in word segmentation. The idea that conditional statistics may be applied to the word segmentation problem was first proposed by Harris (1955) who suggested that individuals working with foreign languages may discover individual word units by counting how many phones can succeed a given string. Across languages, sounds within a word are more predictable than sounds across word boundaries. For instance, the phrase '*prettybaby'* may be segmented as '*pre ttyba by*' or '*pretty baby*'. Due to the phonotactic constraints of the English language, it is more likely that '*pre*' is followed by '*tty*' and '*ba*' is followed by '*by*' than '*tty*' is followed by '*ba*'. To explore whether listeners are sensitive to these conditional statistics, researchers have used artificial languages made up of speech syllables. There are no acoustic cues such as pause or stress on any syllable that may assist in segmenting the words in these

[1] In general, the conditional probability is given by Bayes rule as $p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$ (Jaynes, 2003)

artificial languages. The language streams are constructed such that syllables within a word have higher transitional probabilities compared to syllables across word boundaries.

Sensitivity to these statistics has been evaluated in infants and adults in the seminal studies by Saffran and colleagues (Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996). Twelve consonant vowel (CV) syllables were combined to create six trisyllabic words (*babupu*, *bupada*, *dutaba*, *patubi*, *pidabu*, and *tutibu*). These words were presented in a fluent stream (i.e. no gaps between words) to adults and infants. After a period of exposure to the speech stream, referred to as familiarization, performance was measured through looking times in infants and a behavioural two-alternative forced choice (AFC) task in adults. A measure of learning was obtained when participants discriminated between the words and two different kinds of foil items: non-words (sequences of syllables that had never co-occurred during familiarization, e.g., *tudapu*) or part-words (syllables that co-occurred across word boundaries, e.g., *bupubu*). Both adults and infants could successfully identify the words significantly above chance level (50%), thereby indicating that they were sensitive to the statistical structure of the stream.

Sensitivity to statistical coherence has been replicated using a variety of stimuli. Individuals are sensitive to conditional statistics in auditory, visual, and tactile stimuli (Conway & Christiansen, 2005, 2006; Fiser & Aslin, 2002). Sensitivity to adjacent conditional relationships (e.g., *XY* in A*XY*B) as well as non-adjacent conditional relationships (e.g., *XY* in A*X*B*Y*) have also been documented (Creel, Newport, & Aslin, 2004; Newport & Aslin, 2004). In addition, research has shown that statistical regularities can be learnt in temporally separated sequences (e.g., paradigms similar to Saffran et al., 1996), spatially separated sequences (Fiser & Aslin, 2001), and simultaneously presented sequences (Saffran, 2002).

Some important themes emerge from the vast accumulated literature on learning of conditional statistics. There are significant differences in individual capacities for SL across various modalities. Conway and Christiansen (2005) have shown commonalities but also significant differences in SL in the tactile, auditory, and visual modalities. Individuals were better at learning statistics of temporally presented stimuli in the auditory modality than in the visual and tactile modalities. The authors concluded that different constraints determine SL in a given modality. These findings were further confirmed by Siegelman and Frost (2015) who measured learning of adjacent and non-adjacent conditional statistics in auditory and visual stimuli using five different tasks of SL. They reported the lack of correlation within individuals on performances across the wide range of SL tasks. Siegelman and Frost (2015) suggested that SL is characterized not only by modality specificity but also stimulus specificity. In other words, learners may be consistently sensitive to conditional probabilities for specific types of stimuli or modality, yet consistently insensitive to contingencies in other types of stimuli or modality. As mentioned previously, SL appears to be linked with different cognitive domains such as language comprehension, music appreciation, and perception of speech, hence understanding its underlying mechanisms and constraints are important avenues of future research in various populations (e.g., populations expected to have better SL such as musicians).

1.1.2. Distributional statistics

Distributional statistical learning is thought to play a central role in infants' speech sound discrimination and categorical perception of boundaries. Distributional statistics describe the measures of central tendency of a particular set. For example, learners may be sensitive to measures of central tendency or dispersion (frequency and variability) of exemplars in the input (Maye, Weiss, & Aslin, 2008; Thiessen, 2007). One of the initial works in this area by Maye, Werker, and Gerken (2002) showed that infants' categorical boundaries for phonetic distinctions were based on frequency distribution of speech sounds during exposure or familiarization. Figure 1 shows the example of bimodal and unimodal frequency distribution of stimuli used in their work. When infants are exposed to a bimodal distribution of prototypical |ta| and prototypical |da| (shown in red in Figure 1), they are more likely to discriminate between exemplars from the two categories. In contrast, when infants are exposed to a unimodal distribution of an intermediary sound on the |da| to |ta| continuum (shown in black in Figure 1), they are less likely to discriminate between exemplars from the two categories. Thus, sensitivity to frequency distribution of speech sounds may explain how infants learn the speech sounds of their native language within the first year of life.

Importantly, the role of distributional statistics in language acquisition is not limited to infant speech perception. For instance, distributional cues play a role in learning of syntax by helping learners associate the context in which the exemplars occur (Thiessen, 2007; Thiessen & Yee, 2010).

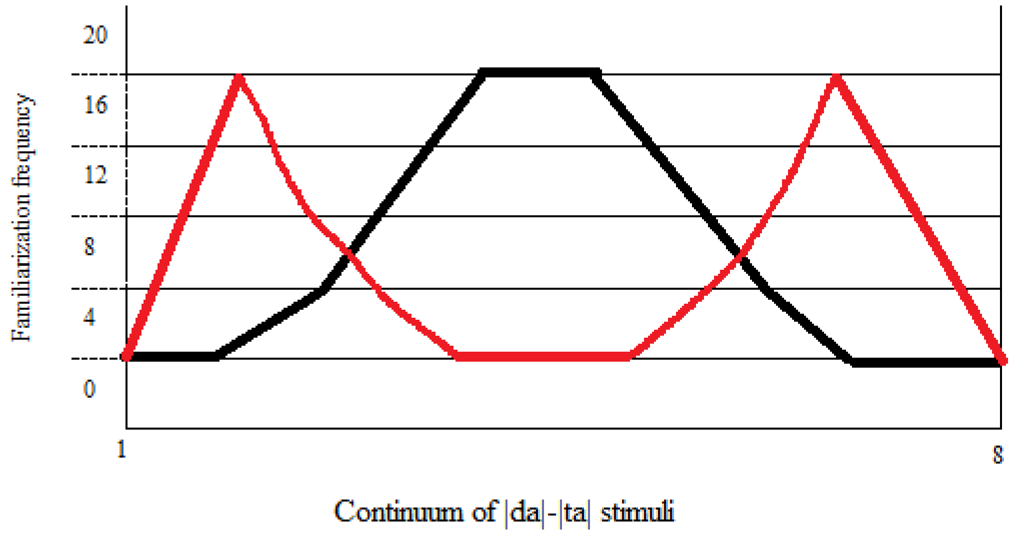

Figure 1: Bimodal (in red) and unimodal (in black) distribution of |da|-|ta| stimuli during familiarization. Redrawn from Maye et al. (2002)

1.1.3. Cue-based statistics

Cue-based SL is a process that learners use to associate perceptible attributes of the input with other non-perceptible attributes and weigh these cues accordingly. Examples of some cues that are not directly perceptible include cues to word boundaries such as pause and stress (Johnson & Jusczyk, 2001). Infants also use phonotactic regularities to segment fluent speech (Saffran & Thiessen, 2003). Importantly, learners generalize their knowledge about cue-based statistics to novel contexts. For instance, first syllable of a word in the English language is usually stressed. Once an infant learns that stress predicts word onsets, they will widely apply the association of stress towards prediction of word onsets even when the stress is on the second syllable of a word (Thiessen & Saffran, 2007). A critical feature of cue-based SL is that the cue, once identified, serves to shape further learning and performance. Attention also plays a role in assisting cue-based SL as learners give more weight, and generalize using perceptually salient cues (Emberson, Liu, & Zevin, 2009). In cue-based SL, learners weigh the use of a cue at least partially as a function of the strength of the probabilistic relationship between the cue and the expected response. Therefore, cue-based SL is not limited to learning of perceptual cues which may be present in conjunction with conditional probability cues. Moreover, different mechanisms/models are used to explain cue-based SL and conditional SL.

1.1.4. Summary

In summary, humans are sensitive to statistical information contained within individual words as well as information that is spread across several stimuli (e.g., exemplars within a category). In this thesis, SL refers to learning of *conditional statistics* to segment a continuous stream of stimuli. Some of the mechanisms that have been proposed to explain how humans are sensitive to one or more of these statistics are discussed in the next section.

1.2. What mechanisms can be used to explain SL?

Thiessen et al. (2013) proposed the ‘extraction and integration approach’ to account for the range of SL phenomena. The process of *extraction* can be defined as “the process of identifying statistically coherent clusters (defined by conditional relations) of perceptual features and storing them in memory as discrete representations (such as word forms)” whereas *integration* is “the process of comparing across those clusters to identify commonalities and the central tendency of the input” (Thiessen et al., 2013, p. 796). The process of extraction can be used to explain sensitivity to conditional statistics but it does not explain distributional or cue-based statistics, which require the ability to generalize from prior experience. Mechanisms behind distributional and cue-based SL are described through the process of integration. In this thesis, SL is discussed in the context of sensitivity to conditional relationships. Thus, models explaining conditional SL are described briefly below.

In order to learn conditional statistics, it is necessary to exploit co-occurrence statistics of adjacent elements for prediction. This mechanism is implemented in boundary finding models where regions of troughs in conditional probabilities (lower transitional probabilities) are inferred to as word boundaries. The computational model of this has been applied in connectionist models/serial recurrent networks (Christiansen, Allen, & Seidenberg, 1998; Elman, 1990). Since likelihood of transitions between elements are stored, boundary findings models predict that as learners become more familiar with a word (e.g., *television*); they should be able to distinguish parts of this word (e.g., *tele*) from random combination of elements. However, a series of experiments by Giroux and Rey (2009) using speech stimuli showed that after exposure to speech stream, participants are actually less able to distinguish parts of word from the random combination of syllables. In other words, if participants were exposed to triplets (‘*patibu*’, ‘*lineri*’ etc.), participants had poor performance when they were asked to distinguish between part of the triplet (e.g., ‘*pati*’ from ‘*patibu*’) and a disyllabic part-word (e.g., ‘*buli*’ made by combining the last syllable of ‘*patibu*’ and the first syllable of ‘*lineri*’).

An alternative approach to explain sensitivity to conditional statistics is based on formation of clusters of statistically related elements or the clustering models. Clustering models propose that a set of statistically coherent clusters are extracted from input and stored in memory as discrete representations. A prominent implementation of clustering is through the process of chunking implemented computationally in PARSER (Perruchet & Vinter, 1998). In PARSER, continuous speech is segmented by creating smaller chunks and these chunks are maintained depending on how often they are re-encountered. In the previous example (learning of *television* vs. *tele*), PARSER predicts that as the exposure increases only words (e.g., *television*) are strengthened and parts of this word (e.g., *tele*) are weakened. This prediction was confirmed by the findings of Giroux and Rey (2009). This model is also used to explain ‘extraction’ as part of the ‘extraction and integration’ framework proposed by Thiessen et al. (2013). Another popular clustering model is Bayesian hypothesis testing (Frank, Goldwater, Griffiths, & Tenenbaum, 2010). The Bayesian models are hypothesis testing models, where a set of hypothesis are formulated about the potential segmentation of the text and then the likelihood of these hypotheses is assessed. Different clustering models invoke very different processes and it is not yet clear which model faithfully simulates human learning processes.

1.3. How is SL measured?

Amongst the various paradigms used to assess SL, two paradigmatic methods to study SL are the embedded triplet tasks and the serial reaction time tasks. These tasks provide a behavioural metric of SL. However, advances in the neurophysiological and neuroimaging techniques have also seen the use of techniques such as electroencephalography (EEG), magnetoencephalography (MEG), near infrared spectroscopy, and functional magnetic resonance imaging (fMRI) to assess SL. The following section briefly describes the behavioural and neurophysiological paradigms assessing SL.

1.3.1. Embedded triplet tasks

In this thesis, the traditional SL tasks are referred to as the embedded triplet tasks consistent with previous SL literature (Arciuli & Simpson, 2011, 2012; Arciuli & von Koss Torkildsen, 2012; Creel et al., 2004; Fiser, 2009). As the name suggests, in the embedded triplets task, the unit of learning is a triplet. Some of the stimuli used to construct triplets in the auditory modality are speech syllables (e.g., *babupu* used by Saffran, Aslin, et al., 1996), musical tones (e.g., *ADB* used by Saffran, Johnson, Aslin, & Newport, 1999), or syllables sung in a particular melodic contour (e.g., *gimysi* used by Schön et al., 2008). Amongst others, stimuli used to construct triplets in the visual modality include non-nameable geometric shapes (Fiser & Aslin, 2002), nameable geometric shapes (e.g., *triangle* used by Abla & Okanoya, 2009), coloured nameable pictures (e.g., *picture of a watch* used by Campbell, Zimerman, Healey, Lee, & Hasher, 2012), real word scenes (e.g., *picture of a waterfall* used by Brady & Oliva, 2008) and cartoon figures (e.g., *pictures of coloured aliens* used by Arciuli & Simpson, 2011). These stimuli are used to construct a familiarization stream or sequence such that stimuli within a triplet have a high transitional probability of co-occurrence (close to 1) whereas stimuli occurring at triplet boundaries have lower transitional probabilities (typically within 0.1 to 0.2).

Figure 2 shows an example of how a familiarization stream may be created. The items can be any of the stimuli mentioned previously. The items are combined to form triplets and the triplets are then randomly concatenated to form familiarization streams.

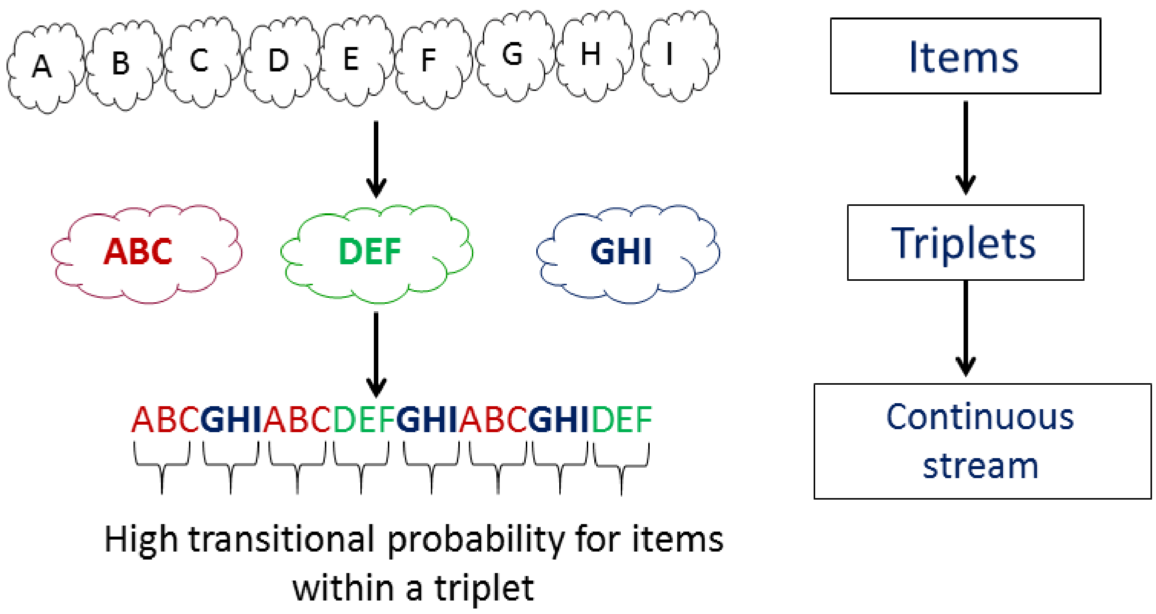

Figure 2: Creation of a hypothetical familiarization stream

Participants may be exposed to the familiarization streams for as little as 2 minutes (Gómez, Bion, & Mehler, 2011) or as long as 42 minutes in case of children and clinical populations (Evans, Saffran, & Robe-Torres, 2009). To encourage participants to pay attention to the familiarization stream without explicitly asking them to track the relationships between stimuli, a cover task is usually employed. Cover tasks may be designed in different ways. For example, participants may be asked to detect an oddball stimulus (e.g., detecting attenuated sounds in Emberson et al., 2009). Alternatively, participants may be asked to detect repeated items such as paradigms where the last item in a triplet is repeated (e.g., *ABCC* in Turk-Browne, Junge, & Scholl, 2005) or one of the items in the triplet is randomly repeated (e.g., *AABC* or *ABBC* or *ABCC* in Arciuli & Simpson, 2012). Familiarization is followed by a test phase where embedded triplets are paired with foil triplets. As mentioned previously, these foil triplets may be novel triplets or part-triplets. At the group level, recognition of the embedded triplets above the chance level of 50% indicates statistically significant learning.

1.3.2. Serial reaction time tasks

Serial reaction time tasks were first adapted by Cleeremans and McClelland (1991) to assess probabilistic sequence learning. In this paradigm, the number of elements in the sequence may vary. Depending on the modality, an auditory or visual cue is linked to a distinct spatially specific motor response. For example, four positions on the computer screen may be linked to four keys on the computer keyboard. The participants' task is to press a button as its corresponding cue is illuminated on the computer screen. Unknown to the participants, the sequence may be a grammatical sequence in which the order of successive stimuli is determined probabilistically. Ungrammatical or violation sequences are also randomly distributed. In these ungrammatical sequences, the order of stimuli is random or has a low transitional probability of co-occurrence. For example, a grammatical sequence A (**1–2**–1–3–2–4), has a probability of 0.85, and an ungrammatical sequence B (**1–2**–4–3-1-4) has a lower probability of occurrence at 0.15. Thus, if the positions '1-2' (highlighted in bold in the two sequences) are illuminated, there is a probability of 0.85 that the next stimulus position is '1' and a probability of 0.15 that the next stimulus position is '4'. An individual is said to have learnt the temporal order statistics if the reaction time for grammatical sequences is significantly shorter than the reaction times for ungrammatical sequences (Hunt & Aslin, 2001; Liu & Liu, 2014). This paradigm has been used to measure SL in auditory as well as visual modalities (Hunt & Aslin, 2001; Kaufman et al., 2010).

Behavioural methods such as the embedded triplet tasks and the serial reaction time tasks remain the most popular methods to study SL. While it remains an open empirical question whether triplet tasks and serial reaction time tasks tap into the same underlying domain-general learning mechanisms, there are some important commonalities as well as differences between the triplet tasks and serial reaction time tasks. For instance, it has been noted that similar brain regions are implicated in SL assessed using the embedded triplet and serial reaction time tasks (Karuza et al., 2013). While both embedded triplet and serial reaction time tasks present rapid sequences of elements, the serial reaction time tasks assess learning through response speed while the embedded triplet tasks rely on a post-exposure test of familiar vs. novel strings. A challenge in using serial reaction time tasks is that when deficits in sequence learning are seen, it may be difficult to ascertain whether these are due to deficits in motor sequencing, rather than in the underlying learning itself (Gamble et al., 2014). Due to the rapid sequencing of element and elicitation of motor responses for each element, obtaining concurrent neurophysiological measures during the serial reaction time task may be challenging. In addition, it has been shown that special populations such as musicians are faster than non-musicians in serial reaction time tasks (on motoric responses) but not necessarily better at SL (Romano Bergstrom, Howard, & Howard, 2012). As the studies in this thesis were designed to investigate the relationship between SL and musical practice using behavioural and neurophysiological measures, the embedded triplet paradigm was utilized.

1.3.3. Neurophysiological and neuroimaging measures

Behavioural studies are limited by their inability to establish the time course of SL. In addition, behavioural tasks cannot distinguish whether extraction of statistical information was achieved via fast, online segmentation or more slow and variable segmentation processes. To understand the brain dynamics as segmentation occurs, we would need to measure brain responses like the N1 and the N400 as the stimuli are presented during an SL task. These brain responses occur within few hundred milliseconds of the stimuli being presented (Kutas & Federmeier, 2011; Näätänen & Picton, 1987). This necessitates the use of techniques like electroencephalography (EEG) and magnetoencephalography (MEG) which are capable of detecting these responses with a suitable temporal resolution such as within milliseconds (Menon, Ford, Lim, Glover, & Pfefferbaum, 1997). Secondly, EEG and MEG provide us with direct evidence about the brain systems involved in online segmentation. Thus, investigating SL using neurophysiological measures in conjunction with behavioural measures helps us to identify neurophysiological correlates of SL and the cortical organization of segmentation systems. Thirdly, it is often difficult to use the same behavioural task with different groups of subjects (e.g., children, adults, clinical populations). For example, tasks investigating SL in infants would have to be modified when used with adults due to the difference in response mode (infants: looking times vs. adults: button press responses). As most EEG paradigms investigating SL do not require explicit responses from participants, the recording of EEG and MEG responses may provide an online measurement of segmentation abilities suitable for listeners of different ages and backgrounds. EEG and MEG also have the potential to inform about the underlying mechanisms involved in SL. In paradigms similar to the embedded triplet paradigm, EEG may be recorded during the familiarization phase or the test phase. Depending on the experimental design, various ERP components have been identified as neurophysiological correlates of SL. In the sections below, results from research using EEG to measure SL and studies identifying brain regions activated during SL tasks are highlighted.

1.3.3.1. Neurophysiological studies

The objective of using EEG to record brain activity during familiarization is to understand the brain dynamics involved in the process of segmentation of continuous stream of events. Sanders, Newport, and Neville (2002) recorded event related potentials (ERPs) as adults listened to six pronounceable trisyllabic non-words presented as continuous speech with no acoustic gaps during familiarization. Following performance at chance level on a subsequent behavioural test, they trained all participants with the six non-words. ERPs were recorded again after training. A word onset effect was observed, that is word onsets (initial syllables) elicited larger N1 responses compared to medial and final positions. High learners based on post-training accuracy scores showed a larger N1 word onset effect.

The authors provided two possible explanations for the word onset effect. Firstly, acoustic differences between the initial and subsequent sounds in a word could result in onset effects. However, a subsequent study by the same group showed that word onset effect was found even when the initial and medial syllables were matched on intensity, duration, and other acoustic characteristics (Sanders & Neville, 2003). Alternatively, it is possible that as listeners become sensitive to transitional probabilities, they direct greater attention to the onset of words which are unpredictable. Increase in amplitude of the N1 response due to attention has been reported previously (Hansen & Hillyard, 1980). This explanation is further supported in another study by Astheimer and Sanders (2011) who found that listeners selectively attend to word onsets that cannot be predicted from the context. Early perceptual processing of information is enhanced when a listener selectively attends to the word onsets resulting in a word onset effect. Taken together, these set of studies illustrate the importance of word onset effect in the process of segmentation.

Cunillera et al. (2009) recorded ERPs as adult participants listened to language streams and random streams. Language streams consisted of four tri-syllabic pseudo-words presented as a statistically coherent stream where transitional probabilities were cues to word boundaries. ERPs recorded during the language and random streams were compared in two minute blocks. A negativity in 350-450 ms region (N400 component) appeared in the second block and then decreased in amplitude in the third and fourth block (described as an inverted U pattern). Further analysis of the first block (0-2 minutes of exposure) showed that a clear N400 component appeared in the second minute of exposure. The N400 was considered as an index of online segmentation. The authors concluded that N400 might reflect the initial protolexical trace that is created after isolating word candidates from the continuous stream.

However, it should be noted that each syllable in this study was presented for 232 ms. Presentation of such short duration stimuli without any inter-stimulus intervals (ISI) could result in overlapping brain responses (Sussman, Steinschneider, Gumenyuk, Grushko, & Lawson, 2008). These methodological limitations must be kept in mind while designing EEG studies to assess SL.

A series of studies by Abla and colleagues used longer stimulus durations to overcome the aforementioned problem. They obtained online measures of segmentation during familiarization using non-verbal auditory and visual stimuli in an embedded triplet paradigm. To examine the word onset effect efficiently, individual stimuli within a triplet were presented for 550 ms without any ISI. Thus, ERPs to initial, middle, and final elements in the triplets could be compared without the limitations of overlap. In the first study, ERPs were recorded as participants listened to three streams consisting of six tone-triplets (Abla, Katahira, & Okanoya, 2008). In a subsequent study, ERPs were recorded as participants watched three streams of six picture-triplets (Abla & Okanoya, 2009). In both studies, participants were divided into three groups (high, middle, low learners) based on performance on the behavioural triplet recognition task.

In the online auditory statistical learning (aSL) study (Abla et al., 2008), a triplet onset effect (larger N1 and N400 elicited to first tone of a triplet) was observed in the high learners which decreased progressively from stream 1 to stream 3. The N400 triplet onset effect appeared only in stream 2 in the middle learners and did not appear at all in the low learners. In the visual statistical learning (vSL) task, a small and sharp N400 triplet onset effect was observed in the first stream of high learners only but no triplet onset effect was found among low learners. However, caution must be exercised while interpreting these results in the vSL paradigm, given the small number of subjects in high learner group (*N*=9). Furthermore, the N400 typically appears as a broad peak of longer duration (Duncan et al., 2009). Therefore, the appearance of a small and sharp N400 triplet onset effect only for high learners needs to be further investigated. Similar to the word onset effect in previous studies, the appearance of the triplet onset effect was attributed to the lower predictability of the first element in a triplet. Distinct ERP effects seen in high learners and low learners demonstrate the sensitivity of ERP measures to the degree of learning. Consistent with Sanders et al. (2002) who used speech stimuli, the studies by Abla and colleagues using non-linguistic stimuli further provide converging evidence that the ERP triplet onset effect can be used as a measure of online segmentation.

François and Schön (2010) recorded ERPs during the test phase of an embedded triplet paradigm. Participants listened to a sung language stream composed of six tri-syllabic words. Each syllable was sung on a distinct tone (e.g., gimysy was sung as C3 D3 F3). To dissociate the effects of music on the learning of words, two types of test phase were designed. On the linguistic test, the words and part-words were spoken with a flat contour. On the musical test, the stimuli were 'midi' sequences of three notes generated with a piano sound (e.g., the notes C3 D3 F3 associated with the word gimysy were played on a piano). A familiarity effect was seen (i.e. larger ERPs for part-words than words) in both musical and linguistic tests. However, the behavioural results of the test phase showed that participants learnt above chance on the linguistic test but not the musical test. The authors interpreted familiarity effect as an ERP correlate of SL.

ERP correlates of SL have also been identified by using modified oddball paradigms. In the auditory domain, a recent study investigated EEG correlates of processing events with different transitional probabilities (Koelsch, Busch, Jentschke, & Rohrmeier, 2016). Stimuli with low transitional probability of co-occurrence were presented as oddballs and elicited a mismatch negativity response (MMN). However, subsequent behavioural testing showed that participants did not learn above chance level of 50%. In the visual domain, Jost, Conway, Purdy, Walk, and Hendricks (2015) investigated EEG correlates of SL by presenting a serial input stream of visual stimuli. A target visual stimulus was preceded by three kinds of predictor visual stimuli. The three predictor visual stimuli differed in how each of them predicted the target stimulus and were either a high probability predictor, a low probability predictor, or a zero probability predictor stimulus. Their findings showed that a P300 component was elicited by the high probability predictor stimulus after sufficient exposure to familiarization stream, over which time the conditional probabilities could be learnt.

Taken together, these studies suggest that the ERP correlates of SL are linked with the type of paradigm used. While collecting EEG data, a number of factors might be kept in mind. First, recording ERPs during the test phase may be influenced by decisional, memory and rehearsal processes (François, Jaillet, Takerkart, & Schön, 2014). Thus, it may not reflect the learning process *per se*. Second, during the test phase participants are required to press a button to indicate which triplet is more familiar to them. This raises the problem of EEG data being contaminated by motor-related ERP activity (Handy, 2005). Third, evoked responses such as the N400 require a minimum of 40-120 artefact free recordings (Duncan et al., 2009). Most embedded triplet paradigms typically contain 36 trials in the test phase where an embedded triplet is paired with a foil triplet on every trial. This raises a concern of insufficient artefact-free ERP recordings during the test phase. Fourth, most ERP studies record data at a number of time points using a large number of electrodes. It is important to control family-wise error rates so as to separate spurious ERP effects from the actual ERP effects. The use of traditional statistical analysis techniques (ANOVAs) is open to user biases of ‘grouping electrodes’ and selecting ‘interesting data’ (Mensen & Khatami, 2013). Lastly, it may be difficult to interpret the ERP correlates of SL when subsequent behavioural tasks show that participants perform at chance or below chance level. This is especially important if a combination of ERP and behavioural methods are used to compare SL performance across populations. It is prudent to use only those paradigms (and stimuli) where learning can be demonstrated above chance level in typical population. This would be helpful in ascertaining whether SL performance is enhanced or diminished when similar tests are administered in special populations (e.g., those with specialized skill sets like musicians or clinical populations). These factors must be kept in mind while designing an ERP paradigm to investigate SL.

1.3.3.2. Neuroimaging studies

There are only a handful of neuroimaging studies that have investigated SL. Results obtained from neuroimaging studies of SL have been mixed, a fact potentially attributable to variation in the behavioural evidence of learning obtained during a scanning session. While neuroimaging techniques may lack the precise temporal resolution obtained through EEG, they have excellent spatial resolution (Josephs, Turner, & Friston, 1997) and can be used to pinpoint brain regions that are activated during SL tasks. This section is included to provide the reader with a background of brain regions activated during SL tasks.

McNealy, Mazziotta, and Dapretto (2006) recorded functional magnetic resonance imaging (fMRI) as participants listened to continuous streams made of four trisyllabic words (not meaningful words in English language) and continuous streams made of random concatenation of syllables. They observed significant bilateral activation in superior temporal gyrus (STG) and transverse temporal gyrus, extending into the supramarginal gyrus (SMG) in the left hemisphere for the statistically coherent stream and not the random stream. However, they found that adult participants were unable to discriminate between statistically coherent and less coherent items during a post-exposure testing phase. The authors proposed that the increased neural activation was the signature of word segmentation, which manifested at a stage before the participants could demonstrate that they recognized the statistically coherent items.

In contrast, Cunillera et al. (2009) conducted a joint ERP-fMRI study of auditory word segmentation and succeeded in obtaining statistically significant behavioural evidence of learning. Using a two-alternative forced choice task, they found that participants could differentiate clusters of statistically coherent syllables from clusters of less coherent syllables. They saw increased activation during the exposure phase in bilateral posterior superior temporal gyrus and the superior part of the ventral premotor cortex (svPMC). These areas identified during word segmentation were consistent with areas activated during phonological processing.

A study by Karuza et al. (2013) attempted to investigate the learning process as it unfolds by changing periods of exposure to a statistically coherent stream. Compared to a random stream of syllables, greater activation was reported in left STG and left inferior frontal gyrus (LIFG/Broca's area). Importantly, Abla and Okanoya (2008) found a similar relationship between the segmentation of continuous tone sequences and activity in inferior frontal cortex. Participants were first trained on isolated tone triplets. Next, these statistically coherent triplets were concatenated in a continuous stream and presented in alternation with random tone sequences. Multichannel near-infrared spectroscopy recordings revealed greater changes in oxy-haemoglobin response localized near the Broca's area for the statistically coherent triplets relative to random sequences. It is interesting to note that speech processing areas and attentional network (IFG) is involved in SL tasks even if the stimuli are non-linguistic (tones).

Examining SL in the visual modality using fMRI, Turk-Browne, Scholl, Chun, and Johnson (2009) offered additional support for the concept of learning without awareness (i.e. before discrimination). They used unfamiliar fonts (referred as glyphs) to create a statistically coherent stream of embedded triplets and a random stream of images. Across the entire exposure phase, they found that participants showed greater activation for statistically coherent relative to random shape sequences in an extensive network of areas including the striatum, medial temporal lobe, lateral occipital cortex (LOC), and ventral occipito-temporal cortex. However, they did not obtain evidence that participants could discriminate statistically coherent shape sequences from less coherent sequences in a subsequent behavioural test phase.

In another study investigating visual statistical learning (vSL) using fMRI Turk-Browne, Scholl, Johnson, and Chun (2010) exposed all participants to a familiarization sequence consisting of 4 paired images (doublets instead of triplets where the first image always predicted the second image). These paired images were randomly sequenced with unpaired images where the first image did not reliably predict the second image. They reported that participants reliably discriminated statistically coherent sequences in the test phase. Interestingly, participants' familiarity ratings of statistically coherent sequences during the test phase were shown to correlate with LIFG activation (an area previously identified as being active during aSL tasks) during the vSL exposure phase.

1.3.4. Summary

Taken together, SL of triplets has been probed using behavioural, neurophysiological, and neuroimaging techniques. Most commonly, N1 and N400 responses have linked to index the process of online segmentation. The above-mentioned findings suggest some overlap in the brain areas involved in the computation of statistical regularities both within and across modalities (e.g., activation of IFG during both auditory and visual SL tasks).

1.4. Musicianship

Music is a multi-modal activity that bolsters amalgamation of auditory, visual, and motor skills. Musicians acquire and continuously practice a variety of complex auditory, motor, and multimodal skills. For instance, during a performance, musicians read musical notations and translate them into motor commands while simultaneously monitoring instrumental output and receiving multisensory feedback (Schlaug, Norton, Overy, & Winner, 2005). Proficiency in music is associated with the acquisition of a set of highly specialized skills (pitch perception, rhythm discrimination, and timber perception to name a few) (Kraus, Skoe, Parbery-Clark, & Ashley, 2009). As a result, the musicians versus non-musician comparison is an ideal model for comparing changes in functional and structural brain plasticity.

Neurophysiological and brain imaging studies have shown differences in the brain structure and function of musicians and non-musicians. For example, professional musicians have been reported to have larger grey matter volume in primary motor and somatosensory areas, premotor areas, anterior superior parietal areas, and in the inferior temporal gyrus bilaterally (Gaser & Schlaug, 2003). In addition, musicians who commenced their training before the age of seven have a larger anterior corpus callosum compared to non-musicians suggesting an increased interhemispheric transfer of auditory information (Schlaug, 2001; for a detailed discussion see Stewart, 2008). The structural differences are found in those brain regions that are closely linked to skills learned that is, auditory and motor skills (i.e. changes seen in primary auditory areas, secondary auditory (association) areas, motor areas) (Schlaug, 2001).

These effects are also observed in children with music training. Schlaug et al. (2005) conducted a longitudinal study to observe brain changes after 1 year of instrumental music training in five- to seven- year old children. Preliminary results showed emerging differences in the grey matter volume and corpus callosum. Data from a cross-sectional study of nine- to eleven-year old children with an average of four years of music training was also reported in the same paper, which showed that children with music training when compared with a control group of untrained children showed larger grey matter volume in sensorimotor cortex and occipital lobes. Moreover, in these children increased activation of STG, posterior inferior (pIFG), and middle frontal gyrus (MFG) was observed using fMRI. Taken together, it is thought that these structural and functional changes underlie the observed differences between musicians and non-musicians in various behavioural auditory and cognitive processing tasks.

In addition, neurophysiological studies (using MEG and EEG) have shown that musicians have better encoding of auditory information. For example, musical training facilitates pitch contour processing not only in music but also in language stimuli (Schön, Magne, & Besson, 2004). When compared to non-musicians, musicians can detect deviant stimuli in longer sequences (up to 8 tones) of complex auditory stimuli (Boh, Herholz, Lappe, & Pantev, 2011). Evidence of enhanced encoding of auditory stimuli is seen at both cortical and subcortical level in musicians (Pantev et al., 1998; Pantev, Roberts, Schulz, Engelien, & Ross, 2001; Tzounopoulos & Kraus, 2009). However, much uncertainty still exists if the enhancements seen in neurophysiological measures are also seen in abilities measures through *behavioural* measures.

One of the most widely investigated behavioural measures in adult musicians pertain to auditory processing tasks. Auditory processing is an umbrella term for “all of the operations executed on peripheral auditory inputs, and which are required for the successful and timely generation of auditory percepts, their resolution, differentiation, and identification” (Phillips, 2002, p. 255). It encompasses spectral, temporal, and binaural processing abilities. Spectral processing involves skills such as frequency resolution, intensity resolution and frequency selectivity (Moore, 2012), while temporal processing involves integration, ordering, sequencing and resolution skills (Shinn, 2003). Binaural processing requires integration, interaction and localization skills which are achieved by using inputs from the two ears (Moore, 1991; Musiek & Chermak, 2013).

Several studies have documented that musicians outperform non-musicians in auditory processing tasks closely related to music (e.g., pitch, timing & timbre) (reviewed in Kraus et al., 2009). For instance, long-term music training has been consistently associated with enhancements in frequency resolution ability in musicians (Fine & Moore, 1993; Kishon-Rabin, Amir, Vexler, & Zaltz, 2001; Micheyl, Delhommeau, Perrot, & Oxenham, 2006). However, there is still uncertainty about the relationship between musical experience and performance in other auditory processing tasks. For example, in some studies musicians performed better than non-musicians on speech-in-noise tasks (amongst others see Parbery-Clark, Skoe, Lam, & Kraus, 2009), temporal resolution tasks (Rammsayer & Altenmüller, 2006) and dichotic listening tasks (Špajdel, Jariabková, & Riečanský, 2007). However, another set of studies showed no differences between musicians and non-musicians on the same tasks (Ishii, Arashiro, & Pereira, 2006; Nelson, Wilson, & Kornhass, 2003; Ruggles, Freyman, & Oxenham, 2014).

The potential reasons for these discrepancies could be due to weak or subtle differences between the groups or different participant recruitment across studies. For instance, when the effects are weak or subtle it is difficult to interpret the differences between the musicians and non-musicians. In a study that showed a musicians' advantage on performance in speech-in-noise tasks, the observed difference in scores between groups was less than 1 dB (Parbery-Clark et al., 2009). Consequently, a study by Ruggles et al. (2014) using similar tasks and participant recruitment criteria as Parbery-Clark et al. (2009) did not see any differences between musicians and non-musicians. The second possible reason for discrepancy between results relates to participant recruitment criteria. For instance, studies that report no group difference on gaps in noise task (e.g., Ishii et al., 2006) and dichotic listening tasks (e.g., Nelson et al., 2003) do not adequately report demographic and music training information to compare results across studies which do report group differences on the same tasks (e.g., Rammsayer & Altenmüller, 2006; Špajdel et al., 2007). Furthermore, there is a lack of research examining performance on a comprehensive battery of auditory processing tasks for a single cohort of musicians and non-musicians. Such a study would enable us to compare the performance of musicians in a systematic manner and circumvent the problem of comparing across studies where participant inclusion criteria differ.

Similar to performance on auditory processing tasks, many studies have compared the performance of musicians on various cognitive processing tasks. Whether or not musical training is associated with enhancements in cognitive processing skills has been widely debated (Degé, Kubicek, & Schwarzer, 2011; Schellenberg, 2011; Schellenberg & Peretz, 2008; Zuk, Benjamin, Kenyon, & Gaab, 2014). For example, Zuk et al. (2014) examined musicians and non-musicians matched for age, gender, IQ, and socioeconomic status. They found better performance in musicians on executive function tasks measuring auditory working memory, cognitive flexibility, and verbal fluency. In contrast, Boebinger et al. (2015) found no differences in auditory working memory and cognitive flexibility between musicians and non-musicians matched for age, gender, IQ, and years of post-secondary education.

It should be noted that both studies (Boebinger et al., 2015; Zuk et al., 2014) recruited highly skilled musicians (those who began training before the age of 9 and had at least 10 years of music training). However, the divergent results between the two studies possibly relate to the reporting of test performances and the choice of statistics. For example, it is possible that Boebinger et al. (2015) did not observe group differences on working memory as they reported a singular value for the forward and backward digit span tests while Zuk et al. (2014) reported only backward digit span results that were significant. Research indicates that performance on the two digit span subtests (forward and backward) should not be combined as they are associated with two distinct memory processes (Reynolds, 1997). This is an important consideration as musical experience may be associated with enhancements of only one memory task (e.g., backward digit span as reported in Clayton et al., 2016; Zuk et al., 2014). Secondly, it is possible that Zuk et al. (2014) report group differences on 4 out of the 6 executive function measures in their study as they used one-tailed *p* values. It is likely that some comparisons between the two groups may not reach statistical significance when a two-tailed test is used (as used by Boebinger et al., 2015). Moreover, very few studies report if corrections for multiple comparisons were applied when repeated measurements are conducted on the same population. For example, studies which do apply multiple corrections show that musicians and non-musician differ on only select measures (Clayton et al., 2016; Helmbold, Rammsayer, & Altenmüller, 2005). These findings further necessitate measuring cognitive and auditory processing in the same group of musicians and non-musicians using a comprehensive test battery. While comparison of intelligence, memory in musicians and non-musicians has been widely reported, an area of cognitive science research that has received scant attention is implicit learning of statistical regularities (SL) in musicians.

1.4.1. SL in individuals with music training

Knowledge of musical structure and expectancies is formed quite early in life (Loui, 2012). It is possible to implicitly acquire musical knowledge and use this knowledge to form expectancies, and extract regularities from continuous events. It has been thought that musicians' prolonged training may be associated with enhancements in the ability to extract statistical information as repeated practice and exposure, primes and sharpens musicians' intuition for performing implicit learning tasks (Rohrmeier & Rebuschat, 2012). Moreover, enhanced encoding of auditory information may also facilitate extraction of statistical information in musicians. For these reasons, it is interesting to compare SL in individuals with and without music training. Table 1 provides a summary of the adult studies comparing SL in musicians and non-musicians. The table includes details such as participant recruitment criteria, stimuli, results, and critical analysis. This table is included to highlight the diversity in approaches adopted to study SL in musicians. Some additional details about these studies are then discussed to highlight how their findings influenced the design of the current study.

Table 1: Summary of studies comparing conditional SL in adult musicians and non-musicians

Studies	Participants	Range of music experience	Stimulus Familiarization	Test phase	Electrophysiology	Results behavioural	Results electrophysiology	Comments
Francois et al., (2011)	16 professional musicians; 20 non-musicians	more than 12 years of musical experience; 3-7 hours of daily practice	5 tri-syllabic words where each syllable was sung on a particular note. Each word repeated 100 times to make a 5.5 min stream.	25 trials 2 AFC task where 2 items (word and part-word) presented on each trial; linguistic test-items were spoken with flat contour; musical test-items were played on piano.	recorded during test phase	Musical test: below chance level performance for both groups. Linguistic test: performance above chance in both groups. No group difference on either musical or linguistic test.	Musical test: familiarity effect at 350-500 ms and 600 ms	Limited number of trials (25) for ERP analysis. Factors such as decision making, muscle activity due to button press response could have affected ERPs. Unclear if the ERP results were reported only for correct trials or all trials. Below chance level performance for most participants may be due to different stimuli being used for familiarization and test phase.
Francois et al., (2014)	13 professional musicians; 13 non-musicians; final analysis 10 participants in each group	same as above	same as above	same as above	recorded during familiarization	Musical test: below chance performance in both groups. Linguistic test: performance above chance in musicians only.	Musicians showed inverted U learning curve for N400; non-musicians showed a linear N400 learning curve.	Difficult to interpret ERP effects as behavioural results were at chance level.
Shook et al., (2013)	15 highly skilled musicians; 15 lower skill musicians	Highly skilled musicians- at least 6 years of music training (mean 11 years); lower skilled musicians-no more than 4 years of music training (mean 2.1 years)	3 two-letter Morse code words. Pure tone (440 Hz) used to represent dots. Testing in statistical condition = transitional probabilities were cues for word boundaries (TaInMeTaMeInTa); Testing in conflicting condition = pauses were added within each word as conflicting cues to word boundaries.	12 trials of 2AFC task where 2 items (word and part-word) were presented on each trial.	no ERPs obtained	Only highly skilled musicians performed above chance level on statistical condition; performance of both groups was below chance level on conflicting condition.	N/A	The authors provided supplementary analysis to show both groups indeed learnt above chance in the conflicting condition but this was not applied to the statistical condition. Thus, it remains unclear if lower skilled musicians may show above chance performance if the new method of analysis was applied.

Skoe at al., (2013)	28 adults in total. It is unclear how many were musically trained.	A range of formal musical (instrumental) instruction, 0-13 years with an average of 4.21±3.9 years.	Musical tone doublets (EC, F#F, DG, G#A) Testing in patterned condition = transitional probabilities were cues to doublet boundaries; Testing in pseudo-random condition = tones were presented randomly with some randomization constraints.	2 AFC task where 2 items - a familiar doublet and a foil item (that had never occurred in patterned condition) were presented on each trial	complex ABR (cABR) recorded during familiarization	Participants (N=28) performed above chance; one-sample t-test results for each group not presented; musically trained individuals outperformed untrained individuals (*p*=0.046)	Smaller cABR for patterned condition than pseudo-random condition. No difference between musically trained and untrained individuals.	It is unclear how many trials were present in the behavioural task. It is difficult to understand the musician group characteristics and results because there was very little information about their musical training. The cABR results for the musicians and non-musicians is not presented separately. Also, it is unclear if both groups performed above chance.
Paraskevopoulos et al., (2012)	15 musicians and 15 non-musicians	Musicians-mean musical training = 16.82; SD = 3.87); Non-musicians-no formal training	Musical tone triplets called standard triplets and oddball triplets. Each block contained 400 standards and 100 oddballs and was 8.08 min in length. Total of 3 blocks presented	36 trials 2 AFC task where each standard triplet was paired with an oddball triplet	MEG recorded during familiarization	There was no difference in performance of two groups. Neither groups performed above chance level.	Larger P50 response (standard vs. oddball) in musicians. However, no difference in MMN amplitude between the groups.	Performance below chance may be due to complexity of pattern along with very small ISI. Larger P50 was attributed to superior auditory encoding in musicians
Romano Bergstrom et al., (2012)	18 video-game players; 18 musicians; 18 controls	Musicians- at least 6 years of training (Mean 10 years); played 6.8 hours/week; Controls 0-9 years of music training (mean 0.6)	high frequency and low frequency triplets Visual serial reaction time task	N/A	no ERPs obtained	Musicians faster than gamers and controls. Musicians also better than controls in overall sequence learning	N/A	It is unclear how accuracy was calculated
Anaya et al., (2016)	24 musicians; 24 non-musicians	Musicians started learning music before 9 years, on an average had 17.3 years of training; played for 18.12 hours/week; non-musicians with little or no music training	4 black squares illuminated to form a pattern visuo-spatial sequence learning task where sequences were presented using 2 grammars: trained grammar and untrained grammar	no ERPs obtained		No difference between musicians and non-musicians on implicit sequence learning	N/A	The groups were not-matched as musicians were older and were more educated than non-musicians. Musicians also had higher scores on non-verbal reasoning tasks.

François and Schön (2011) conducted one of the first studies directly investigating SL in musicians and non-musicians. The authors used a strict inclusionary criterion for the musician group (Table 1). However, musicians did not outperform non-musicians on either the behavioural linguistic test or the musical test assessing aSL. Moreover, both groups performed above chance level on the linguistic test but not the musical test. ERPs for familiar and unfamiliar items were compared such that familiarity effect (larger ERPs for unfamiliar items) was considered a measure of learning. On the musical test only, a significantly larger familiarity effect was observed for only the musicians over the P2 region and over the 350-500 ms region (negative component was interpreted as a mismatch negativity). A fronto-central negativity at 600 ms (interpreted as a delayed N400) was also found in the musician group only. It is difficult to interpret the ERP results for the musical test because learning was not significantly different from chance level in this test. However, the results also suggest differential processing of familiar and unfamiliar items in musicians. Moreover, as mentioned in Table 1 and Section 1.3.3.1, factors such as limited number of trials, decision making, and motor response (when subjects press the button) during the test phase may influence ERPs.

This problem was addressed in a subsequent study that was conducted by the same research group (François et al., 2014). ERPs were recorded during familiarization as musicians and non-musicians listened to the same stimuli as the previous study. Similar to the previous study, performance was at chance level for the musical test in both groups. Contrary to the previous study, only musicians performed above chance level on the linguistic test. ERPs were analysed in four time bins of 80 seconds each and N400 amplitude in each time bin was compared. Non-musicians showed a linear increase in N400 amplitude. In contrast, musicians showed an increase then saturation in N400 amplitude followed by a decrease in amplitude during the fourth time bin (inverted U learning curve). The asymptote and decrease in N400 amplitude was interpreted as evidence that individuals could segment the stream.

Decrease in N400 amplitude showed that as musicians segmented the stream, the sung words became more familiar and the ERPs decreased in amplitude as familiarity increased. The ERP effects indicate that different neural dynamics may underlie segmentation in musicians. However, it is possible that non-musicians may require longer periods of time to show a similar learning curve as musicians.

In contrast, Paraskevopoulos, Kuchenbuch, Herholz, and Pantev (2012), reported no group difference between musicians and non-musicians on both event related fields (ERF) and behavioural measures of SL. However, the authors admitted that performance below chance level in the two groups may be due to the task complexity.

A study by Shook, Marian, Bartolotti, and Schroeder (2013) measured SL in highly skilled musicians and low skilled musicians (details in Table 1) using two-letter Morse code tone-words. The authors reported better performance in highly skilled musicians though it is notable that only highly skilled musicians performed above chance level on the statistical condition. The authors suggested that the SL advantage in musicians may be due to better temporal processing in musicians. However, this claim is yet to be ascertained empirically.

It is difficult to interpret the results by Skoe, Krizman, Spitzer, and Kraus (2013) who measured SL of tone doublets and recorded complex auditory brainstem responses (cABRs) as participants listened to patterned (statistically coherent) and random sequences. Musicians did not have larger cABRs when compared to non-musicians. The authors reported a marginally significant musicians' advantage (p=0.046) in the subsequent 2AFC behavioural task. However, it should be noted that the study reported a wide range of musical instruction in the participants (0-13 years) but did not report the specific criteria used for dividing the groups as musicians and non-musicians.

So far, there is only one study that measured SL in children with music training. François, Chobert, Besson, and Schön (2013) measured aSL using a sung language stream in two groups of eight-year old children in a well-designed longitudinal study. The first group was provided with music training and the second (control) group was provided with painting training. Each training was conducted by a trained teacher (music or painting depending on the group) for two years. The training was conducted for 45 minutes twice a week in the first year and once a week in the second year. SL was measured at three time points: before the training, after the first year of training, and after the second year of training. At each time point, children were exposed to the sung language stream (see Table 1, François & Schön, 2011), but were tested on spoken (linguistic) materials. Behavioural results showed that both groups were at chance level at baseline evaluation. Performance of the musician group continued to improve from baseline. However, performance of the painting group remained at chance level in all the three time points. ERPs recorded during the test phase showed that as opposed to the painting group, a larger familiarity effect was observed in the musician group after two years of training. The longitudinal approach coupled with the pseudorandom assignment of children to the two types of training suggests that music training leads to improvement in SL abilities.

The three studies by François and colleagues used innovative stimuli (sung language) to examine the association between musicianship and SL. However, as noted above, most studies reviewed above report behavioural performance at chance level in the non-musician group. This is surprising, given that a number of studies have previously shown significant learning in typical (non-musician) adults (amongst others see Abla et al., 2008; Saffran et al., 1999; Saffran, Newport, et al., 1996). Using tasks where significant learning has been established in typical (non-musically trained) population, can give us a better understanding if there is any musicians’ advantage on aSL tasks.

As seen in Table 1, to the best of our knowledge, only two studies have investigated implicit sequence learning of visual stimuli in musicians (Anaya, Pisoni, & Kronenberger, 2016; Romano Bergstrom et al., 2012). While Romano Bergstrom et al. (2012) reported better visual sequence learning in musicians than controls, Anaya et al. (2016) report similar performance between the two groups. The mixed pattern of results and the use of unmatched groups (e.g., see Anaya et al, 2016 in Table 1) further raises questions about musicians' advantage in vSL tasks.

1.4.2. Summary

Overall, based on the existing literature, the nature of SL in musicians remains unclear. Below are some salient themes emerging from the SL-musicianship literature:

1. There is inconsistent evidence of musicians' advantage on aSL tasks (marginally significant advantage on the behavioural task and no difference in the cABRs) (e.g., Skoe et al., 2013).
2. Investigation of aSL and vSL in the same individuals can help us answer if the enhancements in SL in musicians may be limited to auditory domain only.
3. Studies in which musician and non-musician groups are matched on age and strict inclusion criteria for the groups have been applied, task complexity may have resulted in chance performance on the behavioural aSL tasks (e.g., François & Schön, 2011; Paraskevopoulos et al., 2012).
4. Recording of ERPs during the test phase may be challenging due to limited number of trials and co-occurrence of motor activity when the participants respond. Obtaining an online measure of learning using ERPs can give us more insights into the brain mechanisms of SL as the process of learning unfolds (e.g., François et al., 2014).

5. It is difficult to interpret group differences in ERP tasks when both groups do not learn above chance level (e.g., François et al., 2014; François & Schön, 2011; Paraskevopoulos et al., 2012).
6. The nature of association between performance on SL tasks and other tasks of auditory and cognitive processing in musicians remains unknown.

The present doctoral research was undertaken to address these gaps in literature. We assessed aSL and vSL in the same group of musicians and non-musicians. The groups were matched on age and education level. The inclusionary criteria for musicians was adopted based on previous research (Boebinger et al., 2015; Parbery-Clark et al., 2009; Ruggles et al., 2014). We used a well-established paradigm (embedded triplet paradigm) and stimuli to study aSL (e.g., tones used by Saffran et al., 1999) and vSL (e.g., cartoon figures used by Arciuli & Simpson, 2011). These paradigms and stimuli have been used to demonstrate significant learning in typical populations (Section 1.3.1). We also included cover tasks to encourage participants to pay attention during familiarization (as highlighted in Section 1.3.1). The problems regarding the design of ERP paradigms which were highlighted in Section 1.3.3.1 were overcome by recording EEG during the familiarization phase to obtain an online measure of SL and using a non-parametric randomization procedure to analyse ERPs (Maris & Oostenveld, 2007). We also assessed performance on auditory or cognitive processing measures to examine the link between these tasks and SL. Bearing in mind these issues, a similar study was designed to examine these questions in children learning music.

1.5. Aims and objectives

The overarching aim of this thesis was to examine the association between musical expertise and SL using behavioural and electrophysiological measures. Specifically, the following questions were addressed:

1. Is music training associated with differences in performance on SL tasks in adults?
2. Is performance on SL tasks linked with performance on auditory or cognitive processing measures?
3. Are there differences in online measures of SL in adult musicians as compared to a control group?
4. How do children with music training perform on behavioural measures of SL as compared to a control group?
5. How do children with music training perform on online measures of SL as compared to a control group?

1.6. Organisation of the thesis

This thesis comprises 4 experimental chapters which are full length manuscripts.

Chapter 2 investigates SL in adult musicians and non-musicians using behavioural tasks. Embedded triplet tasks were used to evaluate SL in the auditory and visual modalities. The experiment examined whether long-term music training is associated with enhancements in performance on aSL and vSL tasks. Furthermore, a comprehensive array of auditory and cognitive processing tasks was used to examine the association between SL and these abilities.

Chapter 3 investigates online measures of both aSL and vSL through EEG recordings in adult musicians and non-musicians. Data collection for Chapters 2 and 3 was undertaken concurrently with the same set of participants. Unlike previous EEG studies in this field (as highlighted in Section 1.3.3.1), we applied a non-parametric permutation testing procedure to analysed ERP effects in SL tasks. This was done to control family-wise errors that arise due to the large number of statistical comparisons involved (evaluating the effect of interest at an extremely large number of sensor-time pairs).

Chapter 4 further explores the association of musical training and performance on SL, auditory processing, and cognitive processing tasks in children. A measure of musical abilities was also evaluated in this population to examine the link between performance on music skills and SL. Both behavioural and online measures of SL in two modalities (aSL and vSL) were examined in children learning music and a control group.

Chapter 5 presents the fourth experiment "*Individual differences in the ability to perceive speech in noise are related to the capacity for statistical learning*". A number of previous studies have examined the link between SL and language processing tasks such as comprehension of syntax, vocabulary learning. However, there is limited literature on how SL might be related to perception of speech in adverse listening conditions (to the best of our knowledge only one such study is by Conway et al., 2010). This study aims to advance our knowledge in this field by evaluating perception of speech by using more everyday listening tasks (speech mixed with background babble).

Chapter 6 summarises the main findings from Chapters 2-5 and Chapter 7 presents the overall discussion of these findings.

The appendices contain: (i) additional analyses, (ii) ethics approval statement, (iii) - (iv) questionnaires used to obtain demographic and musical training information from the participants, and (v) journal publication from the data in this thesis.

1.7. Authors' contributions

PhD candidate (First author)

- Overall design of the studies
- Designing and implementation of SL paradigms used in the current research (coding in presentation software, calibration of signal, trouble-shooting)

- Responsible for implementation of the experimental procedures (recruitment of subjects, testing)
- Responsible for design of analysis procedures (statistical analysis, graphing the results)
- Implementation of EEG analysis in MATLAB under supervision by Dr Ibrahim in chapters 3 and 4
- Interpretation of results in collaboration with supervisors
- Writing of papers

Supervisors' and other co-authors' contribution(s)

- Contributions in overall design of the studies (A/Prof Sharma, A/Prof Arciuli & Prof Demuth)
- Advice on design of SL paradigms (use of embedded triplets, construction of familiarization sequence, nature of cover task, construction of test phase etc.) (A/Prof Arciuli)
- Advice on stimulus optimization for EEG recording (A/Prof Sharma)
- Advice regarding statistical analysis
- ICA data processing for Chapter 4 (Dr Ibrahim)
- Advice on data interpretation
- Data collection for experiment 2 in chapter 5 (Ms Cailyn Furze under the supervision of the PhD candidate)
- Editorial inputs

1.8. References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*(6), 952-964.

Abla, D., & Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: a near-infrared spectroscopy study. *Neuropsychologia, 46*(11), 2787-2795.

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research, 64*(2), 185-190. doi: http://dx.doi.org/10.1016/j.neures.2009.02.013

Anaya, E. M., Pisoni, D. B., & Kronenberger, W. G. (2016). Visual-spatial sequence learning and memory in trained musicians. *Psychology of Music*, 1-17. doi: 10.1177/0305735616638942

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science, 14*(3), 464-473. doi: 10.1111/j.1467-7687.2009.00937.x

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*(2), 286-304. doi: 10.1111/j.1551-6709.2011.01200.x

Arciuli, J., & von Koss Torkildsen, J. (2012). Advancing Our Understanding of the Link between Statistical Learning and Language Acquisition: The Need for Longitudinal Data. *Frontiers in Psychology, 3*, 324. doi: 10.3389/fpsyg.2012.00324

Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia, 49*(12), 3512-3516.

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America, 137*(1), 378-387. doi: 10.1121/1.4904537

Boh, B., Herholz, S. C., Lappe, C., & Pantev, C. (2011). Processing of complex auditory patterns in musicians and nonmusicians. *PloS one, 6*(7), e21458.

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes extracting categorical regularities without conscious intent. *Psychological Science, 19*(7), 678-685.

Campbell, K. L., Zimerman, S., Healey, M. K., Lee, M., & Hasher, L. (2012). Age differences in visual statistical learning. *Psychology and Aging, 27*(3), 650.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*(2-3), 221-268.

Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd, G., Jr. (2016). Executive Function, Visual Attention and the Cocktail Party Problem in Musicians and Non-Musicians. *PloS one, 11*(7), e0157638. doi: 10.1371/journal.pone.0157638

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*(3), 235.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition: International Journal of Cognitive Science*, *114*(3), 356-371. doi: 10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities pitting abstract against stimulus-specific representations. *Psychological Science, 17*(10), 905-912.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *J Exp Psychol Learn Mem Cogn, 30*(5), 1119-1130. doi: 10.1037/0278-7393.30.5.1119

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., . . . Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage, 48*(3), 541-553.

Degé, F., Kubicek, C., & Schwarzer, G. (2011). Music lessons and intelligence: A relation mediated by executive functions. *Music Perception: An Interdisciplinary Journal, 29*(2), 195-201.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., . . . Van Petten, C. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology, 120*(11), 1883-1908.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179-211.

Emberson, L. L., Liu, R., & Zevin, J. D. (2009). *Statistics All the Way Down: How is Statistical Learning Accomplished Using Varying Productions of Novel, Complex Sound Categories?* Paper presented at the 31st annual meeting of the Cognitive Science Society, Amsterdam, the Netherlands.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*(2), 321-335. doi: 10.1044/1092-4388(2009/07-0189)

Fine, P. A., & Moore, B. C. (1993). Frequency analysis and musical ability. *Music Perception: An Interdisciplinary Journal, 11*(1), 39-53.

Fiser, J. (2009). Perceptual learning and representational learning in humans and animals. *Learning & Behavior, 37*(2), 141-153.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499-504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 458.

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

François, C., Jaillet, F., Takerkart, S., & Schön, D. (2014). Faster Sound Stream Segmentation in Musicians than in Nonmusicians. *PloS one, 9*(7), e101340.

François, C., & Schön, D. (2010). Learning of musical and linguistic structures: comparing event-related potentials and behavior. *Neuroreport, 21*(14), 928-932. doi: 10.1097/WNR.0b013e32833ddd5e

François, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex, 21*(10), 2357-2365. doi: 10.1093/cercor/bhr022

Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition: International Journal of Cognitive Science, 117*(2), 107-125.

Gamble, K. R., Cummings Jr, T. J., Lo, S. E., Ghosh, P. T., Howard Jr, J. H., & Howard, D. V. (2014). Implicit sequence learning in people with Parkinson's disease. *Frontiers in human neuroscience, 8*.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience, 23*(27), 9240-9245.

Giroux, I., & Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science, 33*(2), 260-272.

Gómez, D. M., Bion, R. A., & Mehler, J. (2011). The word segmentation process as revealed by click detection. *Language and Cognitive Processes, 26*(2), 212-223.

Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431-436.

Handy, T. C. (2005). *Event-related potentials: A methods handbook*: MIT press.

Hansen, J. C., & Hillyard, S. A. (1980). Endogeneous brain potentials associated with selective auditory attention. *Electroencephalography and clinical neurophysiology, 49*(3), 277-290.
Harris, Z. S. (1955). From Morpheme to Phoneme. *Language: Journal of the Linguistic Society of America, 31*(2), 190-222.
Helmbold, N., Rammsayer, T., & Altenmüller, E. (2005). Differences in primary mental abilities between musicians and nonmusicians. *Journal of Individual Differences, 26*(2), 74-85.
Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General, 130*(4), 658.
Ishii, C., Arashiro, P. M., & Pereira, L. D. (2006). Ordering and temporal resolution in professional singers and in well tuned and out of tune amateur singers. *Pró-Fono Revista de Atualização Científica, 18*(3), 285-292.
Jaynes, E. T. (2003). *Probability theory: The logic of science*: Cambridge university press.
Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*(4), 548-567.
Josephs, O., Turner, R., & Friston, K. (1997). Event-related fMRI. *Human brain mapping, 5*(4), 243-248.
Jost, E., Conway, C. M., Purdy, J. D., Walk, A. M., & Hendricks, M. A. (2015). Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain research, 1597*, 95-107.
Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language, 127*(1), 46-54.
Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition: International Journal of Cognitive Science, 116*(3), 321-340. doi: http://dx.doi.org/10.1016/j.cognition.2010.05.011
Kidd, E., & Arciuli, J. (2016). Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. *Child Development, 87*(1), 184-193.
Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology, 12*(2), 125-144.
Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Sci Rep, 6*.
Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Annals of the New York Academy of Sciences, 1169*(1), 543-557.
Krumhansl, C. L. (2001). *Cognitive foundations of musical pitch*: Oxford University Press.
Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831-843.
Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology, 62*, 621.
Liu, S., & Liu, H. (2014). A Review of Models in Experimental Studies of Implicit Language Learning. *Open Journal of Modern Linguistics, 2014*.
Loui, P. (2012). Statistical learning–what can music tell us. *Rebuschat P, Williams J Statistical Learning and Language Acquisition. Mouton de Gruyter*.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods, 164*(1), 177-190.
Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*(1), 122-134.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition: International Journal of Cognitive Science, 82*(3), B101-B111.
McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *The Journal of Neuroscience*, *26*(29), 7629-7639.
Menon, V., Ford, J. M., Lim, K. O., Glover, G. H., & Pfefferbaum, A. (1997). Combined event-related fMRI and EEG evidence for temporal—parietal cortex activation during target detection. *Neuroreport, 8*(14), 3029-3037.
Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *NeuroImage, 67*, 111-118.
Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*(1), 36-47.
Misyak, J. B., & Christiansen, M. H. (2007). *Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language.* Paper presented at the Proceedings of the 29th Annual Cognitive Science Society.
Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science, 2*(1), 138-153.
Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing*. Netherlands: Brill.
Moore, D. R. (1991). Anatomy and physiology of binaural hearing. *Audiology, 30*(3), 125-134.
Musiek, F. E., & Chermak, G. D. (2013). *Handbook of Central Auditory Processing Disorder, Volume I: Auditory Neuroscience and Diagnosis* (Vol. 1): Plural Publishing.
Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.
Nelson, D. M., Wilson, R. H., & Kornhass, S. (2003). Performance of musicians and nonmusicians on dichotic chords, dichotic CVs, and dichotic digits. *Journal of the American Academy of Audiology*, *14*(10), 536-544.
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127-162.
Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*(6678), 811-814.
Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport*, *12*(1), 169-174.
Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia, 50*(2), 341-349. doi: http://dx.doi.org/10.1016/j.neuropsychologia.2011.12.007
Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing, 30*(6), 653-661.
Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn Sci, 10*(5), 233-238. doi: 10.1016/j.tics.2006.03.006

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*(2), 246-263.

Phillips, D. P. (2002). *Central auditory system and central auditory processing disorders: some conceptual issues.* Paper presented at the Seminars in Hearing.

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal, 24*(1), 37-48.

Reynolds, C. R. (1997). Forward and backward memory span should not be combined for clinical analysis. *Archives of Clinical Neuropsychology, 12*(1), 29-40.

Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science, 4*(4), 525-553.

Romano Bergstrom, J. C., Howard, J. H., & Howard, D. V. (2012). Enhanced implicit sequence learning in college-age video game players and musicians. *Applied Cognitive Psychology, 26*(1), 91-96.

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of Musical Training on Understanding Voiced and Whispered Speech in Noise. *PloS one, 9*(1), e86980.

Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language, 47*(1), 172-196. doi: 10.1006/jmla.2001.2839

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition: International Journal of Cognitive Science*, *70*(1), 27-52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606-621.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*(3), 484.

Saffran, J. R., & Wilson, D. P. (2003). From Syllables to Syntax: Multilevel Statistical Learning by 12-Month-Old Infants. *Infancy, 4*(2), 273-284.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research, 15*(3), 228-240.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature neuroscience, 5*(7), 700-703.

Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology, 102*(3), 283-302.

Schellenberg, E. G., & Peretz, I. (2008). Music, language and cognition: unresolved issues. *Trends in Cognitive Sciences, 12*(2), 45-46.

Schlaug, G. (2001). The brain of musicians. *Annals of the New York Academy of Sciences, 930*(1), 281-299.

Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences, 1060*(1), 219-230.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition: International Journal of Cognitive Science, 106*(2), 975-983.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*(3), 341-349.

Shinn, J. B. (2003). Temporal processing: the basics. *The Hearing Journal, 56*(7), 52.

Shook, A., Marian, V., Bartolotti, J., & Schroeder, S. R. (2013). Musical experience influences statistical learning of a novel language. *The American journal of Psychology*, *126*(1), 95.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105-120.

Skoe, E., Krizman, J., Spitzer, E., & Kraus, N. (2013). The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience, 243*, 104-114. doi: 10.1016/j.neuroscience.2013.03.009

Špajdel, M., Jariabková, K., & Riečanský, I. (2007). The influence of musical experience on lateralisation of auditory processing. *Laterality*, *12*(6), 487-499.

Stewart, L. (2008). Do musicians have different brains? *Clinical Medicine, 8*(3), 304-308.

Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Research, 236*(1), 61-79.

Thiessen, E. D. (2007). The effect of distributional information on children’s use of phonemic contrasts. *Journal of Memory and Language*, *56*(1), 16-34.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin, 139*(4), 792.

Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants’ acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, *3*(1), 73-100.

Thiessen, E. D., & Yee, M. N. (2010). Dogs, Bogs, Labs, and Lads: What Phonemic Generalizations Indicate About the Nature of Children’s Early Word-Form Representations. *Child Development, 81*(4), 1287-1303.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychological Review, 107*(4), 885.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis) similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1131.

Turk-Browne, N. B., Junge, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134*(4), 552-564. doi: 10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21*(10), 1934-1945.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *The Journal of Neuroscience, 30*(33), 11177-11187.

Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron, 62*(4), 463-469.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PloS one, 9*(6), e99868. doi: 10.1371/journal.pone.0099868

This page is intentionally left blank

Chapter 2

Musicians' edge: a comparison of auditory processing, cognitive abilities and statistical learning

Pragati Rao **Mandikal Vasuki**[a, b, c*], Mridula **Sharma**[a, b], Katherine **Demuth**[a, c], Joanne **Arciuli**[b,d]

[a]Department of Linguistics, Australian Hearing Hub, 16 University Avenue, Macquarie University, New South Wales 2109, Australia

[b]The HEARing CRC, 550 Swanston Street, Audiology, Hearing and Speech Sciences, The University of Melbourne, Victoria 3010, Australia

[c] ARC Centre of Excellence in Cognition and its Disorders, Level 3, Australian Hearing Hub, 16 University Avenue, Macquarie University, New South Wales 2109, Australia

[d] Faculty of Health Sciences, University of Sydney, 75 East St, Lidcombe, 1825, Australia

* Corresponding author- Pragati Rao Mandikal Vasuki, Room 602, Level 1, Australian Hearing Hub, Macquarie University, Sydney- 2109, Australia, e- pragati.mandikal-vasuki@mq.edu.au, t- +61 (2) 98504246

Status of the manuscript: Published in Hearing Research [Mandikal Vasuki, P.R., Sharma, M., Demuth, K., & Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research*, 342, 112-123.]

Abstract

It has been hypothesized that musical expertise is associated with enhanced auditory processing and cognitive abilities. Recent research has examined the relationship between musicians' advantage and implicit statistical learning skills. In the present study, we assessed a variety of auditory processing skills, cognitive processing skills, and statistical learning (auditory and visual forms) in age-matched musicians (N=17) and non-musicians (N=18). Musicians had significantly better performance than non-musicians on frequency discrimination, and backward digit span. A key finding was that musicians had better auditory, but not visual, statistical learning than non-musicians. Performance on the statistical learning tasks was not correlated with performance on auditory and cognitive measures. Musicians' superior performance on auditory (but not visual) statistical learning suggests that musical expertise is associated with an enhanced ability to detect statistical regularities in auditory stimuli.

Keywords: Musicians; statistical learning; auditory processing; attention; digit span

Highlights:

- Musicians were better at frequency discrimination and backward digit span
- Musicians had better performance on auditory, but not visual, statistical learning
- Statistical learning ability was independent of other auditory and cognitive skills

2.1 Introduction

Music is a quintessential multisensory activity and musical training involves engagement of multiple neural and cognitive resources. Musicians not only engage in auditory training due to many hours of listening and practising but also multimodal training involving reading and translation of complex symbolic notation into motor activity (Schlaug et al., 2005). Though it is difficult to differentiate abilities that prompt individuals to pursue music training from abilities that may result from music training, some cross-sectional studies comparing musicians with non-musician peers have shown that musicians perform better on certain auditory processing tasks (Fine et al., 1993; Micheyl et al., 2006; Zendel et al., 2012) and tasks of executive function (Bialystok et al., 2009; Zuk et al., 2014). In addition, neurophysiological and brain imaging studies have shown differences in the brain structure and function of musicians and non-musicians. For example, professional musicians have larger grey matter volume in primary motor and somatosensory areas, premotor areas, anterior superior parietal areas, and in the inferior temporal gyrus bilaterally (Gaser et al., 2003). The structural and functional changes associated with musical expertise are in line with an experience-dependent model of neuroplasticity (Münte et al., 2002). In summary, there is widespread interest in the musicians' advantage in various abilities. The current research was designed to explore whether auditory processing, cognitive processing, and implicit learning of statistical regularities – statistical learning – differ as a function of musical expertise.

2.1.1 Auditory processing and musical expertise

Some of the most widely investigated abilities associated with musical expertise pertain to auditory processing. Auditory processing is an umbrella term including spectral and temporal processing. Additionally, measures of temporal processing may include envelope processing and fine structure processing. The tests used to measure auditory processing include tests of frequency discrimination, discrimination of iterated rippled noise, detection of amplitude modulation, detection of gaps in noise, dichotic listening tests, and perception of speech in noise.

When considering the relationship between musical expertise and auditory processing, two issues come into play. First, as far as we are aware, no previous study has compared the performance of musicians and non-musicians on auditory processing tasks that address both spectral and temporal processing abilities. Second, the extent to which musical expertise is associated with superior performance on auditory processing tasks remains a controversial issue. On one hand, research has consistently shown that musicians have better frequency discrimination than non-musicians (Kishon-Rabin et al., 2001; Micheyl et al., 2006). On the other hand, contradictory evidence exists for musicians' superior auditory skills for gap detection (Ishii et al., 2006; Rammsayer et al., 2006; Zendel et al., 2012), perception of speech in noise (Parbery-Clark et al., 2009; Ruggles et al., 2014), dichotic listening tests (Nelson et al., 2003; Špajdel et al., 2007), and other temporal processing skills (Iliadou et al., 2014; Ishii et al., 2006). Given these inconsistencies in the literature, we incorporated a comprehensive battery of auditory processing tasks in the current study.

2.1.2 Cognitive abilities and musical expertise

Musical expertise might be associated with some aspects of cognitive processing. For instance, it has been reported that adults and children who have undertaken music training have better working memory as measured through digit span and non-word span (Lee et al., 2007). Better performance on executive function measures such as verbal fluency, design fluency and backwards digit span for musicians has also been reported (Zuk et al., 2014). Still, there have been some ambivalent results as to whether or not musical expertise is associated with enhanced cognitive abilities. Using a large battery of tasks assessing cognitive skills such as verbal comprehension, word fluency, mental rotation, closure, perceptual speed, reasoning, and verbal memory, Brandler et al. (2003) found significant group differences in only two tasks – verbal memory and reasoning. Similar results were reported in another study where musicians were found to have better performance in only two out of the thirteen primary cognitive abilities tested – flexibility of closure and perceptual speed (Helmbold et al., 2005). Additionally, it is unclear whether enhancements are seen only in the auditory modality, such as in auditory attention tasks (Strait et al., 2011), or also in the visual modality, such as divided visual attention tasks (Rodrigues et al., 2007). Given these gaps in the literature, we incorporated a battery of cognitive processing tasks in the current study including a task that assessed both visual and auditory attention.

2.1.3 Statistical learning and musical expertise

A growing area of interest is musicians' ability to learn statistical regularities implicitly, known as statistical learning (SL). SL was described in a seminal study by Saffran et al. (1996). They showed that participants are able to extract statistical regularities from a continuous stream of individually presented stimuli using information about transitional probabilities. SL has been shown in auditory (aSL) and visual (vSL) modalities. It is thought that SL ability may contribute to key mental activities including musical appreciation, object recognition, and language acquisition (Arciuli & von Koss Torkildsen, 2012; Rohrmeier et al., 2012).

Similar to language, music is highly structured and listeners are able to extract regularities from music (François et al., 2010). Whilst being unaware of the complex patterns of music, it is possible to implicitly acquire musical knowledge and use this implicit knowledge to form expectancies, and extract regularities from continuous events. Heightened sensitivity to these statistical regularities in continuous speech or non-speech streams may be partly explained by the shared and overlapping cortical regions for music and language (OPERA hypothesis; Patel, 2010, 2011). It could also be argued that musical competence, which is acquired through repeated practise and exposure, primes and sharpens musicians' intuition for performing implicit learning tasks (Rohrmeier et al., 2012). In addition, musicians have may have enhanced processing of auditory stimuli (for a detailed review see François & Schön, 2014). For these reasons, it is interesting to study SL in musicians.

Using neurophysiological measures such as electroencephalography and magnetoencephalography, musicians have been shown to have enhancements in neurophysiological indices (such as N100 or N400) in auditory tasks involving the extraction of distributional cues (François, Jaillet, et al., 2014; François et al., 2011; Paraskevopoulos et al., 2012; Schön et al., 2011). To date, only two studies have demonstrated an advantage for adult musicians in aSL using behavioural indices (Shook et al., 2013 using morse code; Skoe et al., 2013 using tone doublets). A report of improved SL in a longitudinal study of 8- year old children learning music as opposed to a control painting group suggests that there may be a causal link between musical training and SL (François et al., 2013). Although musical expertise has been associated with improved skills in the visual domain, such as enhanced recognition of visual patterns, also known as design learning (Jakobson et al., 2008), an investigation of musicians' vSL has not been undertaken previously.

Any demonstrable musicians' advantage in SL raises further questions as to whether such an advantage is accompanied by advantages in auditory processing or other cognitive skills. Though not directly investigated, enhanced statistical learning of morse code in musicians was attributed to enhanced temporal encoding and/or cognitive skills in musicians (Shook et al., 2013). However, as far as we are aware, this has not been investigated empirically. We used an array of auditory processing tasks and cognitive processing tasks as well as measures of both auditory and visual SL to explore these questions.

2.1.4 The current study

The primary aims of this research were to ascertain whether musicians and non-musicians perform differently on: a) tests of auditory processing, b) tests of cognition, and c) tests of SL (aSL and vSL). We hypothesized that musical expertise would be associated with better performance on at least some of the auditory processing and cognition measures. We also hypothesized that musicians might outperform non-musicians with regard to aSL but we were not sure what to expect with regard to vSL. Moreover, we were unsure whether performance on SL tasks would be related to performance on the auditory and cognitive tasks.

2.2 Methods

2.2.1 Participants

Musicians were defined as adults who started to learn/practise music before the age of 9 years and had at least 10 years of music playing/singing experience. This criterion is based on previous studies with similar populations (Ruggles et al., 2014; Strait et al., 2010). All musicians reported that they still actively practised music. Non-musicians had less than 3 years of musical experience. Eighteen musicians (5 males) and 22 non-musicians (5 males) participated in the study. There was no significant difference in the ages of the musicians (Mdn=28.0) and non-musicians (Mdn=25.0) as assessed by a Mann-Whitney U test [$U=163$, $p=0.35$]. All participants were right handed as assessed by the Edinburgh Handedness

Inventory (Oldfield, 1971) and had normal or corrected to normal vision. At the time of data collection, four of the non-musicians were enrolled in the final year of a postgraduate program. Details about participants' music education, instruments played, and educational background are in Table 1.

The participants were recruited through advertisements distributed via group emails or announcement on Facebook community pages. All participants lived in the greater Sydney metropolitan area and had obtained an undergraduate degree. The study was approved by the Macquarie University Human Participants Ethics Committee and written consent was received from all participants. All participants were paid $40 for their participation.

Table 1 Details of all participants' musical and educational background

No	Age of onset (years)	Years of training	Instrument/s	Highest degree obtained
M1	5	48	Piano, recorder, hackbrett	Post graduate
M2	7	28	Piano, flute	Post graduate
M3	4	18	Piano, guitar, vocals, drums, bass guitar	Graduate
M4	7	13	Piano	Post graduate
M5	9	16	Guitar	Graduate
M6	4	22	Piano, violin	Doctorate
M7	9	20	Vocals, guitar	Post graduate
M8	5	20	Piano, vocals (soprano)	Post graduate
M9	5	16	Piano, alto saxophone, flute, vocals	Graduate
M10	7	17	Piano, vocals, guitar, trombone	Post graduate
M11	8	25	Piano, vocals	Doctorate
M12	5	19	Guitar	Post graduate
M13	6	10	Piano, vocals	Graduate
M14	7	15	Piano	Graduate
M15	9	12	Piano, violin	Graduate
M16	7	16	Violin, bass guitar	Graduate
M17	6	52	Piano, percussion	Doctorate
M18	4	25	Piano, bongo, percussion	Graduate
NM1	-	0	-	Post graduate
NM2	6	2	Piano	Doctorate
NM3	11	2	Trombone	Graduate
NM4	-	0	-	Graduate

NM5	-	0	-	Graduate
NM6	-	0	-	Graduate
NM7	10	3	Piano	Doctorate
NM8	-	0	-	Post graduate
NM9	-	0	-	Post graduate
NM10	-	0	-	Graduate
NM11	-	0	-	Graduate
NM12	-	0	-	Graduate
NM13	-	0	-	Doctorate
NM14	-	0	-	Graduate
NM15	-	0	-	Graduate
NM16	-	0	-	Graduate
NM17	-	0	-	Graduate
NM18	-	0	-	Post graduate
NM19	-	0	-	Post graduate
NM20	-	0	-	Graduate
NM21	-	0	-	Graduate
NM22	-	0	-	Graduate

2.2.2 Tests

All participants completed behavioural testing in a sound treated booth. Participants' hearing was screened at 15 dB HL (all octave frequencies from 0.5 to 8 kHz). Additionally, distortion product otoacoustic emissions (DPOAEs) and contralateral acoustic reflexes were present at clinically normal levels in all participants.

2.2.2.1 Auditory processing

The auditory processing tests were the dichotic digits test (3 pairs) (Strouse et al., 1999), gaps in noise test (Baker et al., 2008), frequency discrimination of 1 kHz tones, threshold for discrimination of iterated rippled noise (Peter et al., 2014), threshold for detection of 4 Hz and 64 Hz amplitude modulation (Peter et al., 2014), and the listening in spatialized noise sentence test (LiSN-S, Cameron et al., 2007). As test materials and scoring procedures have been published previously, only brief details of the administration and scoring procedures are mentioned in Table 2.

2.2.2.2 Cognition (memory, inhibition, and attention)

Working memory capacity was evaluated using forwards and backwards digit span subtests from the clinical evaluation of language fundamentals, 4^{th} edition (CELF-IV; Semel et al., 2006). A Stroop colour word test was used to evaluate inhibition and selective attention. The integrated visual auditory continuous performance test was used to measure sustained attention in auditory and visual modalities (IVA-CPT, Turner et al., 1995). A brief description of procedures for the tests is given in Table 2.

Table 2 Details of various tests used with a brief description of procedure

Measures	**Tests**	**Procedure**
Auditory processing	Frequency discrimination test (1 kHz) (Peter et al., 2014)	*Stimuli:* 1 kHz pure tone served as standard stimulus. The variable (target) stimuli were generated with frequencies ranging from 1001 Hz to 1050 Hz in steps of 1 Hz. All stimuli were 500 ms duration with a ramp of 20 ms. *Procedure:* Thresholds were estimated based on a 3 AFC procedure with a 2-down 1-up tracking method, estimating the 70.7% correct point on the psychometric function (Levitt, 1971). The target signal frequency was reduced after 2 correct responses and was increased after 1 incorrect response. *Response and scoring:* The participant's task was to identify the interval containing the different signal. The step size was initially 5 Hz and was reduced to 1 Hz after two reversals. The arithmetic mean of the last three reversals in a block of 6 was taken as threshold. Log transformation was applied to the thresholds.
	Threshold for discrimination of iterated rippled noise (Peter et al., 2014)	*Stimuli:* The iterated ripple noise (IRN) stimuli were generated using MATLAB 7 using the add-original configuration (IRNO) method described by Yost (1996). The standard stimulus was white noise with zero iteration. The variable IRNO stimuli were created by adding 10 ms delayed copies of white noise with the original noise. The process was repeated 8 times. IRNO were generated at different gain factors (g) that is, attenuation of the delayed repetition relative to the original noise in order to obtain versions of the stimuli with different pitch strengths. The g ranged from 0 to 0.2 in steps of 0.01. All the stimuli were 500 ms in duration with 30 ms rise and fall times. *Procedure:* Thresholds were estimated based on a 3 AFC procedure with a 2-down 1-up tracking method, estimating the 70.7% correct point on psychometric function (Levitt, 1971). In this

		procedure, the g of the target signal was reduced after 2 correct responses, and was increased after 1 incorrect response. *Response and scoring:* Participants indicated which of the three stimuli had a pitch percept. The step size for the variable stimuli was initially 0.02 and was reduced to 0.01 after two reversals. The arithmetic mean of the last six reversals in a block of 12 was taken as threshold.
	Gaps in noise test (Baker et al., 2008)	*Stimuli:* White noise with duration of 500 ms with ramp of 20 ms was used as the standard stimulus. White noise of 500 ms duration with varying durations of silence inserted in the centre were used as variable stimuli. *Procedure:* Thresholds were estimated based on a 3 AFC procedure with a 2-down 1-up tracking method, estimating the 70.7% correct point on psychometric function (Levitt, 1971). In this procedure, duration of silence in the target signal was reduced after 2 correct responses, and was increased after 1 incorrect response. *Response and scoring:* Participants indicated which of the three stimuli contained a gap. The step size was initially 3 ms and was reduced to 1 ms after two reversals. The arithmetic mean of the last 3 reversals in a block of 6 was taken as threshold.
	Detection of amplitude modulation (Peter et al., 2014)	*Stimuli:* The standard stimulus was a white noise low pass filtered at 20,000 Hz. The standard stimulus was amplitude modulated with varying modulation depths to create variable stimuli. The modulation frequencies used were 4 and 64 Hz. *Procedure:* 3 AFC with a 2-down 1-up tracking method, estimating the 70.7% correct point on psychometric function (Levitt, 1971) was employed to estimate threshold. In this procedure, the depth of modulation in the target signal was reduced after 2 correct responses, and was increased after 1 incorrect response. *Response and scoring:* Participants were asked to indicate which of the three stimuli was not constant. The amplitude modulation thresholds were based on the modulation depth in decibels

		($20*\log_{10}(m)$). The step size was initially 4 dB and was reduced to 2 dB after two reversals. The arithmetic mean of the last three reversals in a block of 6 was taken as threshold.
	Dichotic digits test (3 pairs) (Strouse et al., 1999)	*Stimuli:* Over the course of 25 trials, three pairs of digits were presented dichotically (total of 6 digits per trial). *Response and scoring:* After each trial, participants were asked to repeat as many of the 6 digits as they could. They were not asked to report digits from each ear separately. The recalled digits were scored according to which ear they were presented and percentage correct scores were obtained. The right ear advantage was calculated by subtracting left ear scores from right ear scores.
	Listening in Spatialised noise (sentences test) (Cameron et al., 2007)	*Stimuli:* The commercially available Listening in Spatialized Noise Sentence test (LiSN-S) was used. Target sentences were mixed with distracter discourse and presented at a simulated 0 degree azimuth. Two subtests were administered: a) target and distractor discourse were spoken by the same speaker i.e. same voice-0° and b) target and distractor were spoken by different speakers i.e. different voices-0°. *Response and scoring:* Participants were asked to repeat target sentences and ignore the distractor discourse. The level of the target sentence was adjusted adaptively by the software to estimate the speech reception threshold (SRT) for the two subtests. The levels were adjusted in 4 dB steps until the first reversal in performance was recorded and in 2 dB steps thereafter.
Cognition	Forwards and backwards digit span test (Semel et al., 2006)	*Stimuli:* Digits were presented through headphones at the rate of one digit per second. *Response and scoring:* Participants were asked to recall the digits in same order (forward span) or reverse order (backward span). Age referenced scaled scores were obtained from raw scores using procedures described in CELF-IV (Australian) norms.

	Stroop Colour Word test	*Stimuli:* Computerised test which consisted of three subtasks: naming the ink colour in which the symbol 'X' is printed (Stroop I), reading the colour names printed in black ink (Stroop II), and naming the ink colour of the printed words in which ink color and the word differ (e.g., the word 'red' printed in green ink, Stroop III). The test was implemented in Presentation software (www.neurobs.com). *Response and scoring:* Four response buttons with names of colours printed in black ink were used. Participants pressed a button depending on whether they were asked to identify the ink colour or colour name. Reaction times were obtained for each trial. Mean reaction times were calculated for all the subtasks. Mean colour word interference score was calculated by subtracting the average time needed to complete the first two subtasks from the time needed to complete the third subtask (interference score=Stroop III - [(Stroop I + Stroop II)/2]) (Van der Elst et al., 2006).
	Integrated visual auditory continuous performance test (IVA-CPT) (Turner et al., 1995)	*Stimuli:* Computerised test where 500 trials of visual and auditory '1's and '2's were presented pseudorandomly requiring a shift between the two modalities. *Response and scoring:* Over a span of thirteen minutes, the participants were asked to click the mouse every time they saw or heard '1'. The number '2' was to be ignored. Sustained Auditory and Visual attention quotients are automatically generated by the reporting component of the IVA-CPT software, and represent age- and gender-matched population norms.

2.2.2.3 Statistical learning

SL was investigated unimodally in the auditory and visual domains. The separate aSL and vSL tasks were designed to be as similar as possible, using the embedded triplet paradigm with a familiarization phase followed by a separate test phase. The aSL and vSL tasks were adapted from Abla et al. (2008), Abla et al. (2009), and Arciuli et al. (2011).

2.2.2.3.1 Stimuli used for familiarization

The stimuli used for creating embedded triplets for the aSL familiarization phase were same as those described in previous studies (Abla et al., 2008; Saffran et al., 1999). Eleven pure tones within the same octave (starting at middle C or 261.6 Hz within a chromatic set) were generated using MATLAB (R2013 a). Tones are labelled according to their musical notation and are depicted in Figure 1. The tones were 550 ms in duration (25 ms rise time and 25 ms fall time). Three tones were combined in succession to form a triplet. All participants were exposed to a familiarization stream containing 6 triplets (ADB, DFE, GG#A, FCF#, D#ED, CC#D) similar to those described by Saffran et al. (1999).

For the vSL task, stimuli comprised the 11 cartoons (described as 'aliens'), used by (Arciuli et al., 2011; Arciuli & Simpson, 2012a). Stimuli are depicted in Figure 1. Cartoons were scaled such that the maximum height and width of all cartoons were equal. Each cartoon was presented for 550 ms (similar to tones in the aSL task) against a black background. Three cartoon figures were combined in succession to form a triplet. As with the aSL task, all participants were exposed to 6 visual triplets during familiarization.

Triplet	aSL	vSL
Triplet 1	ADB	
Triplet 2	DFE	
Triplet 3	GG#A	
Triplet 4	FCF#	
Triplet 5	D#ED	
Triplet 6	CC#D	

Figure 1: Triplets used for the aSL and vSL tasks. Stimuli were presented unimodally.

2.2.2.3.2 Creation of familiarization stream:

Irrespective of the modality of the SL task, the familiarization streams were constructed in the following manner. Triplets were concatenated in a pseudorandom order to form a continuous stream of stimuli with no gaps between triplets. Thus, triplets were 'embedded' in a continuous stream. This stream consisted of 40 repetitions of each triplet. The familiarization stream was made up of 240 triplets (40 presentations of each of the 6 triplets) and was about 7 minutes long. Three familiarization streams were created with different pseudorandom ordering of triplets. The triplets were concatenated with two randomization constraints as described in previous studies (Arciuli et al., 2011; Arciuli & Simpson, 2012b): a) no repeated triplets were allowed and b) no repeated triplet pairs were allowed. The statistical structure for the aSL and vSL familiarization streams was identical. Statistical cues were the only indicator of triplet boundaries as there were no discernible discontinuities at the triplet boundaries.

Statistical cues to the presence of triplets included transitional probabilities (TP) within triplets and across triplet boundaries (Figure 2). For example, the within-triplet TP for CC#D in Figure 2 was calculated as the mean of TPs of the doublets 'CC#' and 'C#D'. For the aSL task and the vSL task, the TPs within triplets for all three streams ranged from 0.25 to 1 (mean 0.625). In contrast, the TPs across triplet boundaries were 0.04-0.3 (mean 0.11).

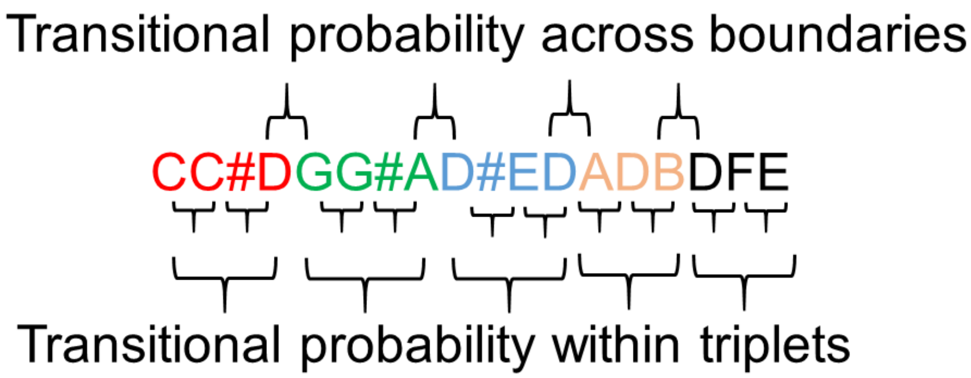

Figure 2: Calculation of transitional probabilities (TP) for a representational aSL familiarization stream.

In the aSL familiarization stream, the mean frequencies for the initial, middle, and final tones across all triplets were 341 Hz (SD = 66), 321 Hz (SD = 57), and 370 Hz (SD = 82), respectively. There was no significant difference between the frequencies of the tones at each position within the tone triplets [F (2, 10) = 1.48, p = 0.28]. The mean frequencies of the tones within a triplet were as follows- ADB= 409.2 Hz, DFE= 324.2 Hz, GG#A=415.8 Hz, FCF#= 326.9 Hz, D#ED=311.5 Hz, and CC#D=277.5 Hz. There was no significant difference between mean pitch interval within versus across triplets (3.1 vs 4.9 half tones in average).

2.2.2.3.3 Creation of the test phase

For each of the SL tasks, the test phase included 36 two alternative forced choice trials (2 AFC) (Abla et al., 2009; Saffran et al., 1999). For each trial during the test phase, participants were presented with an embedded triplet and a novel triplet (counter-balanced order of presentation). The novel triplets were made up from the same individual stimuli but had never occurred together as an embedded triplet in the familiarization stream. The novel triplets were formed according to those reported by Saffran et al. (1999). Each individual tone or cartoon within a triplet was presented using the same presentation time (550 ms) as had been used during familiarization. For each trial, the presentation of the embedded versus the novel triplet was separated by a 1-second gap. After the presentation of both triplets, a new screen appeared which prompted participants to identify which of the two triplets had appeared previously (during familiarization).

Before beginning the test phase, four practice trials were performed to ensure that participants waited for presentation of both triplets before responding. The practice trial triplets did not occur during the familiarization or test phases. This was done to avoid interference between the practice and test trials. Verbal feedback was given during the practice trials. No time constraints were imposed during the practice or test trials.

2.2.3 Procedure

Data were collected over two sessions. Hearing screening, auditory processing tests, and tests of cognition were administered in session 1. The aSL and vSL tests were administered in session 2. Stimuli for assessments in session 1 were presented using a laptop PC through headphones (Telephonics TDH-39) at 50 dB HL via an Interacoustics AC-40 clinical audiometer. The aSL and vSL task stimuli were delivered through Presentation software (Version 16.5, www.neurobs.com). The aSL stimuli were presented via Etymotic ER 3A insert ear phones (at 70 dB SPL). The visual stimuli were presented on a 17-inch CRT monitor placed at 1 metre distance from the participant.

2.2.3.1 Familiarization phase for SL

Participants were exposed to auditory and visual stimulus streams during which they performed an oddball detection task. This task served as a cover task to ensure attentiveness. During the familiarization phase of the aSL task, participants were asked to press a button whenever they heard a pure tone with a frequency of 1319 Hz (not used in any embedded or novel triplets). During the vSL familiarization phase, they were asked to press a button whenever they saw a particular alien figure (not used in any embedded or novel triplets). To ensure that learning was implicit, participants were not given any instructions about the nature of the embedded triplets within the familiarization stream and were not told to learn or remember anything. There were 40 presentations of the oddball stimulus within each familiarization stream for each modality. The oddball stimulus was randomly presented at the end of a triplet.

Half the participants undertook the familiarization phase of the aSL task before the familiarization phase of the vSL task and the order was reversed for the other participants. The entire familiarization phase for both aSL and vSL lasted about 50 minutes (2 modalities X 3 streams X 7 minutes). Breaks of up to 5 minutes were given between presentations of the streams. Participants were informed about the upcoming test phase only after completion of the familiarization phases for both aSL and vSL tasks.

2.2.3.2 Test phase

The aSL and vSL test phases were administered in the same order as the familiarization phases. For instance, a participant who was first exposed to the aSL familiarization phase completed the aSL test phase task first. The duration of the test phase for each SL task was 5 minutes.

2.2.4 Data analysis

The data from 1 musician and 4 non-musicians were excluded for reasons including active middle ear pathology (1 non-musician) and a score of 0 on the IVA- CPT indicative of attention deficits. Hereafter, any description of the results does not include data for these 5 participants. Thus, data from 17 musicians (4 males, median age 26 years) and 18 non-musicians (4 males, median age 25.5 years) are reported. A Mann-Whitney U test showed no significant difference between the ages of the retained musicians and non-musicians [$U = 143.5$, $p = 0.76$].

Mann-Whitney U tests were used to compare performance on tests of auditory and cognitive processing between the two groups as the data were not normally distributed. Due to multiple comparisons for tests of auditory processing and cognition, we considered *p* values <0.01 as significant. A Wilcoxon signed-rank test was used to assess if musical expertise was associated with modality specific enhancements in the attention tasks. Normality for performance on the SL tasks was confirmed using Shapiro-Wilk tests. Pearson’s correlations were used to explore associations between SL and tasks of auditory or cognitive processing. All statistical analyses were performed using SPSS version 20 software.

2.3 Results

2.3.1 Auditory processing

The mean, median, standard deviation, and results from the Mann-Whitney U test for all tests of auditory processing are shown in Table 3. A significant group difference was found for only one auditory processing measure: frequency discrimination. Musicians had better (lower) frequency discrimination thresholds than non-musicians.

2.3.2 Cognition

Table 4 presents the mean, median, standard deviation and statistics for all the measures of cognition. A significant group difference was found only for backwards digit span, with musicians exhibiting significantly higher scores than non-musicians. There were no group differences on tests of inhibition (Stroop task) and sustained attention. A Wilcoxon signed-rank test showed that the two groups performed at a similar level for auditory and visual attention tasks (musicians: $Z = -0.55$, $p = 0.6$; non-musicians: $Z = -1.45$, $p = 0.15$). There was no difference between musicians and non-musicians on either auditory or visual attention tasks. Additionally, modality specific enhancements in attention were not observed for musicians.

Table 3: Means, medians, SDs and effect sizes for performance on tests of auditory processing. Significant results are in bold font.

Test	Musicians			Non-musicians			Statistics		
	Mean	Median	SD	Mean	Median	SD	U	*p*	r
Frequency discrimination test (log transformed)	**0.7**	**0.7**	**0.2**	**1.1**	**1**	**0.8**	**53.5**	**0.001**	**-0.6**
Threshold for discrimination of iterated rippled noise: IRN-to-noise gain ratio	0.07	0.05	0.02	0.07	0.07	0.03	123.5	0.3	-0.2
Gaps in noise test (ms)	2.7	2.6	0.4	2.6	2.6	0.4	133.5	0.5	-0.1
Detection of amplitude modulation (4 Hz)- Modulation depth (20 $\log_{10}$ m)	-23.7	-22.7	2.5	-23.3	-24	4.9	149.5	0.9	-0.01
Detection of amplitude modulation (64 Hz)- Modulation depth (20 $\log_{10}$ m)	-14.5	-14.7	3.3	-13.6	-12.7	2.9	121.5	0.3	-0.2
Dichotic digits test – right ear scores (%)	94.8	94.7	4.3	94.2	94.7	4.2	143.5	0.8	-0.4
Dichotic digits test – left ear scores (%)	94.4	94.7	5.3	89.9	89.3	6.5	82	0.02	-0.05
Dichotic digits test (right ear advantage)	0.4	0	5.9	4.4	4.7	5.1	94.5	0.05	-0.3
Listening in Spatialized noise different voices-0° (SRT in dB)	-7.8	-7.4	3.5	-7.9	-7.7	3.1	144.5	0.8	-0.04
Listening in Spatialized noise same voices-0° (SRT in dB)	-2.9	-2.5	2.2	-3.1	-2.5	2.2	139.5	0.7	-0.07

Table 4: Means, medians, SDs and effect sizes for performance on tests of cognition. Significant results are in bold font.

Test	Musicians			Non-musicians			Statistics		
	Mean	Median	SD	Mean	Median	SD	U	*p*	r
Forwards digit span	10.9	11	1.8	9.5	10	2.8	110.5	0.2	-0.24
Backwards digit span	**12.2**	**13**	**2.5**	**9.5**	**9.5**	**2.1**	**57**	**0.001**	**-0.53**
Stroop colour word interference score (ms)	134.3	110.8	80.7	154.5	104	163.9	145	0.8	-0.04
Sustained auditory attention quotient	108.8	110	18.3	103.4	110	30.1	147.5	0.9	-0.03
Sustained visual attention quotient	106.1	111	18.0	96.9	105.5	28.7	105	0.1	-0.27

2.3.3 Statistical learning

Data from the oddball detection tasks during familiarization were analysed to determine the percentage of successfully identified stimuli. Each participant scored above 80% in the aSL and vSL oddball detection cover tasks. The percentage of correctly identified embedded triplets during the test phase was recorded for musicians and non-musicians. Consistent with previous SL studies (Arciuli & Simpson, 2012a; Conway et al., 2010; Stevens et al., 2015), participants who scored outside the range mean ± 2 SD were excluded from further analyses. This resulted in the exclusion of 4 participants – one non-musician for an excessively low score on the aSL task, one musician for an excessively high score on the vSL, and two non-musicians for excessively low scores on the vSL task. Thus, 34 participants were retained for aSL (17 musicians) and 32 participants for vSL (16 musicians).

Figure 3 depicts the percentage of correctly identified embedded triplets for retained participants on the aSL and vSL tasks. We performed item analyses to ensure that responding was consistent across the six embedded triplets presented during the test phase. For example, the transitional probabilities were higher within some triplets than others. Results showed that responding was consistent across triplets for both groups. Normal distributions of performance on the aSL and vSL tasks was confirmed for both groups using the Shapiro-Wilk test (all *p*s >0.05). One sample t-tests revealed that musicians and non-musicians performed significantly above chance on the aSL task [musicians $t(16) = 14.9$, $p < 0.001$, $d = 3.62$; non-musicians $t(16) = 6.1$, $p < 0.01$, $d = 1.49$]. Likewise, performance on the vSL task was significantly above chance for both groups [musicians $t(15) = 2.7$, $p < 0.05$, $d = 0.67$; non-musicians $t(15) = 2.9$, $p < 0.05$, $d = 0.73$].

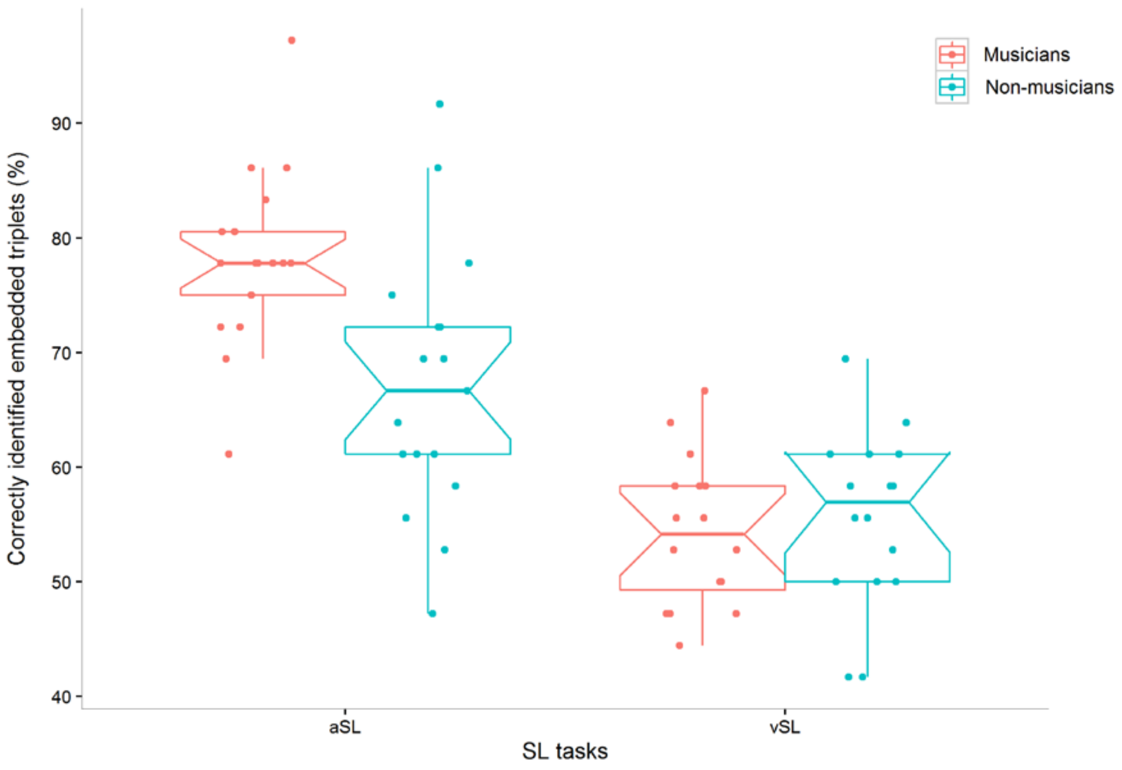

Figure 3: Box and whisker plots showing the percentage of embedded triplets correctly identified by musicians and non-musicians on the aSL and vSL tasks. The box denotes 75th to 25th percentile (interquartile range). The upper whisker represents 75th percentile + 1.5 x IQR and the lower whisker represents 25th percentile -1.5 x IQR. 50% indicates chance performance.

We conducted a 2x2 ANOVA to compare performance on the two SL tasks (modality: aSL and vSL) across the two groups (musicians and non-musicians). There was a significant main effect of modality [F (1, 29) =61.57, p<0.005, partial η^2=0.68], and of group [F (1, 29) =5.71, p<0.05, partial η^2=0.16], and a significant interaction between group and modality [F (1, 29) = 9.87, p<0.01, partial η^2=0.25]. Pairwise comparisons using t-tests with Bonferroni correction revealed that musicians significantly outperformed non-musicians on the aSL task [t (30) = 3.29, p<0.05, d= 1.13] but not on the vSL task [t (30) = -0.49, p>0.05, d= 0.17].

Scores on the aSL and vSL tasks were not correlated (all participants: r= -0.02, p>0.05; musicians: r= 0.37, p>0.05; non-musicians: r=-0.15. p>0.05). For the musicians, there were no significant associations between SL performance and age of onset of music training (aSL: r=0.47, $p>0.05$; vSL: r=0.34, $p>0.05$) or between SL and years of music training (aSL: r=-0.23, $p>0.05$; vSL: r=-0.1, $p>0.05$).

2.3.4 Relationship between statistical learning and measures of auditory processing and cognition

Combining the data for both groups, we performed a Pearson's correlational analysis to explore the relationships between SL and measures of auditory processing and cognition. Scores for aSL were not correlated with any auditory processing or cognitive measures except frequency discrimination. There was a moderate negative correlation between performance on aSL and frequency discrimination thresholds (r=-0.43, $p=0.01$). To tease apart the influence of musical expertise in this association, Pearson's r was calculated separately for each group. The results revealed no significant relationships (musicians: r= 0.35, $p=0.17$; non-musicians: r=-0.45, $p=0.07$)[1]. vSL was not correlated with any of the measures of auditory processing or cognition. Based on the results of these correlational analyses and to avoid the risk of over-fitting, a follow up multiple-regression analysis was not conducted.

2.4 Discussion

Our findings indicate that musicians performed better than non-musicians in one auditory and one cognitive task: frequency discrimination and backward digit span. A key finding was that musicians outperformed non-musicians in identifying embedded triplets in the aSL task but not the vSL task. Performance on the SL tasks was not correlated with performance on the auditory and cognitive processing tasks.

[1] We also confirmed this using a non-parametric Spearman's correlation. Performance on aSL and frequency discrimination was not correlated for either group (musicians: $r_s=0.35$, $p=0.17$; non-musicians: $r_s=-0.4$, $p=0.11$)

2.4.1 Auditory processing

The musicians had lower frequency discrimination thresholds than the non-musicians and this difference was statistically significant. Enhanced frequency discrimination performance of musicians has been documented previously (Micheyl et al., 2006; Spiegel et al., 1984). Similar to the current study, Micheyl et al. (2006) found that musicians with classical music training of more than 10 years had lower frequency discrimination thresholds than non-musicians. Kishon-Rabin et al. (2001) found that classical musicians had smaller frequency discrimination thresholds than contemporary (e.g. jazz, modern) musicians. It is noteworthy that the musicians in our study were also trained in classical music. Thus, smaller frequency discrimination thresholds may be due to emphasis on correct tuning during classical training (Micheyl et al., 2006).

We measured frequency discrimination at 1 kHz which has been suggested to be dominated by the use of temporal information that is, a phase locking mechanism (Moore, 2012). Musicians have been shown to have more precise temporal phase locking at the brainstem level as indexed by the frequency following response (FFR) (Lee et al., 2009) . Better phase locking could be one of the reasons for finer frequency discrimination in musicians. Furthermore, it has been suggested that enhanced frequency discrimination could be due to better short-term memory representation or increased attention (Kishon-Rabin et al., 2001; Tervaniemi et al., 2005). However, for the musician group in our study, there was no correlation between frequency discrimination and performance on working memory and attention tasks. It is possible that the relationship between frequency discrimination and cognitive measures may be better probed by using different tests of attention and working memory than those used in the current study.

A right ear advantage (REA) was observed for both groups in the dichotic digits task. The difference in REA scores between the groups was not statistically significant. Musicians and non-musicians also had similar performance on the gap detection test. These findings are in agreement with previous literature comparing musicians and non-musicians on dichotic listening tasks involving digits (Nelson et al., 2003), and gaps in noise tasks (Ishii et al., 2006; Monteiro et al., 2010). Taken together, these findings suggest that superior performance of musicians may be limited to specific auditory tasks.

There were no group differences for detection of sinusoidal amplitude modulation of noise and discrimination of iterated rippled noise. Though a previous study (Lee et al., 2009) reported stronger encoding of temporal envelope cues in the brainstem of musicians using an electrophysiological technique (FFR), we found no evidence of this using behavioural measures. Enhancements in electrophysiological measures (such as increase in the amplitude of FFR) might not necessarily translate to enhancements in behavioural measures (Bidelman et al., 2011). Further research using more comparable behavioural and electrophysiological paradigms is needed to provide information about temporal envelope processing in musicians.

Musicians and non-musicians had similar speech perception in noise scores measured through the LiSN-S test. This is in line with three studies that also investigated speech perception in noise using a variety of tests (e.g. Boebinger et al., 2015; Fuller et al., 2014; Ruggles et al., 2014). In contrast, in other studies of speech perception, musicians outperformed non-musicians (Parbery-Clark et al., 2009; Parbery-Clark et al., 2011; Parbery-Clark et al., 2012). These differences could be due to the different tests used across the studies. However, Ruggles et al. (2014), using the same tests and participant criteria as used by Parbery-Clark et al. (2009), found no significant advantage of being a musician on speech perception in noise tests (Experiment 2). Additionally, findings from Parbery-Clark et al. (2009) suggested that group differences may be observed only when the masking tasks are difficult, for example when the target and maskers are co-located. We administered two subtests of the LISN-S test, where the target speech and masker were co-located (different voices-0° and same voice-0°). However, there was no difference between the groups even when maskers were co-located and had no fundamental frequency cues (the hardest condition, i.e. same voice-0°). A recent study (Clayton et al., 2016) also found no significant group differences when the masker and target speech were co-located. Further research involving stricter criteria for selection of participants (for example, recruitment of only professional musicians) and assessment of performance on a wider range of speech in noise tests may assist in understanding the relationship between speech perception in noise and musical experience. Overall, our results suggest that musical expertise might be associated with enhancements in only a subset of auditory perceptual skills.

2.4.2 Cognition

The current findings add to the converging evidence for musicians' superior performance on working memory tasks (Chan et al., 1998; Franklin et al., 2008; Ho et al., 2003). It has been suggested that musicians allocate more brain resources as the working memory load increases. This was evidenced by larger blood oxygenation-level dependent (BOLD) signal measured with functional magnetic resonance imaging (fMRI) in musicians compared to non-musicians during an n-back working memory task (Pallesen et al., 2010).

Interestingly, we found that musicians had significantly better scores on backward digit span but not on forward digit span. It has been suggested that distinct cognitive processes are tapped by forward and backward digit span tests and this may be why we observed that musicians outperformed non-musicians on backward but not forward digit span. For instance, the difference in performance on the forward and backward digit span tests may be explained by two theoretical approaches – the complexity view and the representational view (Rosen et al., 1997).

According to the complexity view (Rosen et al., 1997), backward recall involves considerable attentional demands, with manipulation of digits held in short-term memory, and is thus considered to be a part of executive function processes tapping into working memory (Rosen et al., 1997). The second approach, the representational view, argues that backward recall may involve specific visuospatial processing where items are represented in a spatial array for easier reversal (Li et al., 1995; St Clair-Thompson et al., 2013). During the course of training and practice, musicians are required to learn melody and memorize sequences of notes either by learning through ear or through visual memory of music notations. A reasonable hypothesis to explain these group differences is that musicians improve their working memory storage through practice (consistent with the complexity view). Alternatively, musicians may employ visual strategies (such as visualizing each digit) to assist them with backward recall (consistent with the representational view). Anecdotally, a few musicians in our study reported using visual strategies during the backward recall task.

In the current study, musicians did not have enhanced performance on the visual Stroop task which is consistent with evidence that musical experience may be linked with specific components of executive function such as backwards recall but not inhibition (Boebinger et al., 2015; Clayton et al., 2016; Zuk et al., 2014). It is noteworthy that (Bialystok et al. (2009)) using an *auditory* Stroop task demonstrated enhanced inhibition in musicians. Further studies are required to understand the influences of modality on executive function tasks.

Although it has been suggested that musical training contributes to enhanced auditory but not visual attention (Strait et al., 2010), our results indicated no enhancements in either visual or auditory sustained attention. This could be attributed to the different tasks and types of attention measured in the two studies. Strait and colleagues measured *alertness* in one modality at a time by comparing reaction times in the presence or absence of a variable delay cue. In contrast, our study used the IVA-CPT test (Table 2) which measured *sustained* attention in auditory and visual modalities concurrently by accounting for accuracy as well as reaction time in the presence of a distracting stimulus. Overall, the mixed pattern of results observed between this and other studies suggests that the relationship between musical expertise and general cognitive abilities is complex and needs further investigation.

2.4.3 Statistical learning

To the best of our knowledge, this is the first study showing behavioural differences in aSL between musicians and non-musicians using an embedded triplet task comprising tones. In previous studies that have assessed aSL using both electrophysiological and behavioural measures in musicians and non-musicians, group differences were observed only for electrophysiological measures (François et al., 2011; Paraskevopoulos et al., 2012). However, it should be noted that behavioural test results in the latter study did not differ significantly from chance for either group. In our study, both groups showed a mean aSL that was significantly better than chance, however, musicians outperformed non-musicians.

Though the current research design cannot address the question of whether musical training causes enhancement of aSL in adults, a randomized longitudinal training study of 8-year old children suggested a causal relationship between musical training and aSL (François et al., 2013). The enhanced aSL for musicians in our study may be attributed to the fact that Western tonal music is based on a strong system of regularities between musical events, such as notes and chords (Tillmann et al., 2000). Our findings suggest that long-term musical expertise could help form associations between successive stimuli and group them into distinctive units (triplets in our case) purely based on the transitional probabilities. Although the triplets we used did not follow the rules of any standard music composition, it could be argued that musicians' pre-existing knowledge regarding relative pitches of the Western scale could make it easier to detect statistical regularities amongst tones using the Western scale

(Loui, 2012). Using a new, unfamiliar Bohlen-Pierce scale, Loui and colleagues reported no differences in performance of musically trained and untrained subjects for melody recognition and rule generalization (Loui et al., 2010). However, it should be noted that the musician group in that study had less musical training than the participants in our study (range of 5 to 14 years with an average of 9.6 years). In contrast, the musicians in our study had at least 10 years' experience of singing or playing music (range of 10 to 52 years with an average of 21.6 years). A systematic study using professional musicians or musicians with at least 10 years of musical training might help to clarify if the duration of the musical training is relevant in statistical learning of unfamiliar musical scales.

Notwithstanding the fact that some previous research has shown that musical expertise is associated with an enhanced visuospatial iconic representation (Gromko et al., 1998), and visual imagery (Neuhoff et al., 2002), musicians and non-musicians in the current study had comparable performance on the vSL task. This finding could be taken as support for an experience-dependent modality specific enhancement of SL that is, the musicians' advantage in SL was only seen in one modality (aSL) and not the other modality (vSL). It is possible that more efficient usage of visuospatial perception and imagery in musicians (Brochard et al., 2004) might translate into enhancements in vSL measured using stimuli that contains spatial regularities. Further research is needed to explore these possibilities.

Better performance in the aSL task than in the vSL task by both groups suggests that SL proceeds differently across the visual and auditory modalities (Conway et al., 2005, 2009). Additionally, there was no significant correlation of the scores for the aSL and vSL tasks. This is in line with recent research (Frost et al., 2015; Siegelman et al., 2015). Given that musical practice involves the integration of multi-modal stimuli, future studies may further explore the role of musical expertise in SL using an auditory-visual SL task.

2.4.4 Relationship between statistical learning and measures of auditory processing and cognition

Shook et al. (2013) postulated that enhanced SL in musicians may be due to their enhanced temporal processing of auditory information as reflected in temporal discrimination, gap detection and rhythm perception tasks (Rammsayer et al., 2006). However, in our study, performance on both SL tasks was independent of performance on any auditory processing task for both musicians and non-musicians. Future studies could use an aSL task involving stimuli with closer frequency separation to advance our knowledge about the relationship between aSL and frequency discrimination in musicians. In addition, further investigation using a variety of auditory processing tasks involving music perception (such as chord perception, melody discrimination) could help shed light on what strategies musicians use for extracting distributional cues.

Our finding that SL performance was not related to cognitive ability is in line with previous studies where SL was not associated with measures of verbal reasoning, intelligence and working memory (Kaufman et al., 2010). Siegelman et al. (2015) also reported that there was no relationship between scores on SL (aSL and vSL tasks using the embedded triplet paradigm) and a large battery of cognitive tests (working memory, verbal working memory, rapid automatized naming, and switch task). These findings point towards the independence of SL from other general cognitive abilities. However, it should be noted that the relationship between SL and cognitive abilities such as working memory is complex and warrants further investigation. For instance, Arciuli and Simpson (2011), argued that implicit rather than explicit working memory tasks might reveal a relationship with SL. Furthermore, SL performance measured through 2AFC trials may involve additional processes such as decision making (François, Jaillet, et al., 2014). Future studies may investigate the relationship between SL and cognitive abilities by incorporating concurrent cognitive tasks as SL takes place and using a wide battery of SL tasks (including SL tasks that do not rely on 2AFC trials).

It has been hypothesized that SL operates automatically with little or no dependence on executive attentional resources (Turk-Browne et al., 2005). Indeed, our results showed that SL performance was not correlated with measures of attention and inhibition. For both SL tasks used in our study, the participants were engaged in a cover task during familiarization that required them to process the stimuli in a different way (to detect oddballs rather than to extract statistical information per se). It is notable that the participants were successful in recognising embedded triplets in both the aSL and vSL tasks despite this competing demand that meant they were actively trying to process the stimuli in another way. As long as participants attend to the relevant stimuli, aSL and vSL may indeed operate automatically.

As music consists of a variety of temporal patterns, future research may investigate the link between aSL and other aspects of auditory processing, such as rhythm perception. Finally, a longitudinal study involving music training and measurement of SL, along with a comprehensive auditory processing test battery could shed some light into causality and the mechanisms underlying enhanced aSL.

2.4.5 Summary

In summary, using a set of auditory processing, cognitive, and statistical learning measures, we found that long-term musical expertise is associated with better performance on frequency discrimination and backward digit span tasks. We also present the first empirical evidence of better aSL, but not vSL for musicians than for non-musicians as evaluated using the embedded triplet paradigm. Performance on SL tasks was not correlated with performance on any other auditory processing or cognition measures.

2.5 Acknowledgements

We would like to thank the editor Brian Moore and three anonymous reviewers for their helpful comments. This work was funded by the Australia Awards Scholarships and the HEARing CRC, established and supported under the Cooperative Research Centres Program – an Australian Government Initiative. This work was also supported by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021). Demuth was supported by a Laureate Fellowship granted by the Australian Research Council (FL130100014). Arciuli was supported by a Future Fellowship granted by the Australian Research Council (FT130101570).

2.6 References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*(6), 952-964.

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research, 64*(2), 185-190. doi: http://dx.doi.org/10.1016/j.neures.2009.02.013

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science, 14*(3), 464-473. doi: 10.1111/j.1467-7687.2009.00937.x

Arciuli, J., & Simpson, I. C. (2012a). Statistical learning is lasting and consistent over time. *Neuroscience Letters, 517*(2), 133-135. doi: 10.1016/j.neulet.2012.04.045

Arciuli, J., & Simpson, I. C. (2012b). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*(2), 286-304. doi: 10.1111/j.1551-6709.2011.01200.x

Arciuli, J., & von Koss Torkildsen, J. (2012). Advancing Our Understanding of the Link between Statistical Learning and Language Acquisition: The Need for Longitudinal Data. *Frontiers in Psychology, 3*, 324. doi: 10.3389/fpsyg.2012.00324

Baker, R. J., Jayewardene, D., Sayle, C., & Saeed, S. (2008). Failure to find asymmetry in auditory gap detection. *Laterality, 13*(1), 1-21.

Bialystok, E., & DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 565-574.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition, 77*(1), 1-10.

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America, 137*(1), 378-387. doi: 10.1121/1.4904537

Brandler, S., & Rammsayer, T. H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychology of Music, 31*(2), 123-138.

Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition, 54*(2), 103-109. doi: http://dx.doi.org/10.1016/S0278-2626(03)00264-1

Cameron, S., & Dillon, H. (2007). Development of the listening in spatialized noise-sentences test (LISN-S). *Ear and Hearing, 28*(2), 196-211.

Chan, A. S., Ho, Y.-C., & Cheung, M.-C. (1998). Music training improves verbal memory. *Nature, 396*(6707), 128-128.

Clayton, K. K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A. D., & Kidd, G., Jr. (2016). Executive Function, Visual Attention and the Cocktail Party Problem in Musicians and Non-Musicians. *PloS one, 11*(7), e0157638. doi: 10.1371/journal.pone.0157638

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition: International Journal of Cognitive Science, 114*(3), 356-371. doi: 10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24.

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21*(4), 561-580.

Fine, P. A., & Moore, B. C. (1993). Frequency analysis and musical ability. *Music Perception: An Interdisciplinary Journal*, *11*(1), 39-53.

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

François, C., Jaillet, F., Takerkart, S., & Schön, D. (2014). Faster Sound Stream Segmentation in Musicians than in Nonmusicians. *PloS one*, *9*(7), e101340.

François, C., & Schön, D. (2010). Learning of musical and linguistic structures: comparing event-related potentials and behavior. *Neuroreport, 21*(14), 928-932. doi: 10.1097/WNR.0b013e32833ddd5e

François, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex, 21*(10), 2357-2365. doi: 10.1093/cercor/bhr022

François, C., & Schön, D. (2014). Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: the role of musical practice. *Hearing Research, 308*, 122-128.

Franklin, M. S., Moore, K. S., Yip, C.-Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*(3), 353-365.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117-125.

Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*, *8*, 179. doi: 10.3389/fnins.2014.00179

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience, 23*(27), 9240-9245.

Gromko, J. E., & Poorman, A. S. (1998). Does perceptual-motor performance enhance perception of patterned art music? *Musicae Scientiae*, *2*(2), 157-170.

Helmbold, N., Rammsayer, T., & Altenmüller, E. (2005). Differences in primary mental abilities between musicians and nonmusicians. *Journal of Individual Differences*, *26*(2), 74-85.

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology, 17*(3), 439-450.

Iliadou, V. V., Bamiou, D. E., Chermak, G. D., & Nimatoudis, I. (2014). Comparison of two tests of auditory temporal resolution in children with central auditory processing disorder, adults with psychosis, and adult professional musicians. *International Journal of Audiology, 53*(8), 507-513. doi: 10.3109/14992027.2014.900576

Ishii, C., Arashiro, P. M., & Pereira, L. D. (2006). Ordering and temporal resolution in professional singers and in well tuned and out of tune amateur singers. *Pró-Fono Revista de Atualização Científica, 18*(3), 285-292.

Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception: An Interdisciplinary Journal, 26*(1), 41-55.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition: International Journal of Cognitive Science, 116*(3), 321-340. doi: http://dx.doi.org/10.1016/j.cognition.2010.05.011

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology, 12*(2), 125-144.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *The Journal of Neuroscience, 29*(18), 5832-5840.

Lee, Y.-s., Lu, M.-j., & Ko, H.-p. (2007). Effects of skill training on working memory capacity. *Learning and Instruction, 17*(3), 336-344.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(2B), 467-477.

Li, S.-C., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 837.

Loui, P. (2012). Statistical learning–What can music tell us. In Rebuschat P & W. J (Eds.), *Statistical Learning and Language Acquisition*. Mouton de Gruyter.

Loui, P., Wessel, D. L., & Kam, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception: An Interdisciplinary Journal, 27*(5), 377-388.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*(1), 36-47.

Monteiro, R. A. M., Nascimento, F. M., Soares, C. D., & Ferreira, M. I. D. d. C. (2010). Temporal resolution abilities in musicians and no musicians violinists. *Arquivos Internacionais de Otorrinolaringologia, 14*(3), 302-308.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing*. Netherlands: Brill.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience, 3*(6), 473-478.

Nelson, D. M., Wilson, R. H., & Kornhass, S. (2003). Performance of musicians and nonmusicians on dichotic chords, dichotic CVs, and dichotic digits. *Journal of the American Academy of Audiology, 14*(10), 536-544.

Neuhoff, J. G., Knight, R., & Wayand, J. (2002). *Pitch change, sonification, and musical expertise: Which way is up?* Paper presented at the International Conference on Auditory Display.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PloS one, 5*(6), e11120.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia, 50*(2), 341-349. doi: http://dx.doi.org/10.1016/j.neuropsychologia.2011.12.007

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing, 30*(6), 653-661.

Parbery-Clark, A., Strait, D. L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia, 49*(12), 3338-3345.

Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience, 219*, 111-119.

Patel, A. D. (2010). *Music, language, and the brain*: Oxford university press.

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

Peter, V., Wong, K., Narne, V. K., Sharma, M., Purdy, S. C., & McMahon, C. (2014). Assessing spectral and temporal processing in children and adults using Temporal Modulation Transfer Function (TMTF), Iterated Ripple Noise (IRN) perception, and Spectral Ripple Discrimination (SRD). *Journal of American Academy of Audiology, 25*(2), 1-9.

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal, 24*(1), 37-48.

Rodrigues, A. C., Guerra, L. B., & Loureiro, M. A. (2007). *Visual attention in musicians and non-musicians: a comparative study*. Paper presented at the Proceedings of the 3rd International Conference on Interdisciplinary Musicology.

Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science, 4*(4), 525-553.

Rosen, V. M., & Engle, R. W. (1997). Forward and backward serial recall. *Intelligence, 25*(1), 37-47.

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of Musical Training on Understanding Voiced and Whispered Speech in Noise. *PloS one, 9*(1), e86980.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition: International Journal of Cognitive Science, 70*(1), 27-52.

Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences, 1060*(1), 219-230.

Schön, D., & François, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology, 2*, 167. doi: 10.3389/fpsyg.2011.00167

Semel, E. M., Wiig, E. H., & Secord, W. (2006). *CELF 4: Clinical Evaluation of Language Fundamentals*: Pearson: Psychological Corporation.

Shook, A., Marian, V., Bartolotti, J., & Schroeder, S. R. (2013). Musical experience influences statistical learning of a novel language. *The American journal of Psychology, 126*(1), 95.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105-120.

Skoe, E., Krizman, J., Spitzer, E., & Kraus, N. (2013). The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience, 243*, 104-114. doi: 10.1016/j.neuroscience.2013.03.009

Špajdel, M., Jariabková, K., & Riečanský, I. (2007). The influence of musical experience on lateralisation of auditory processing. *Laterality, 12*(6), 487-499.

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America, 76*(6), 1690-1695.

St Clair-Thompson, H. L., & Allen, R. J. (2013). Are forward and backward recall the same? A dual-task study of digit recall. *Memory & Cognition, 41*(4), 519-532.

Stevens, D. J., Arciuli, J., & Anderson, D. I. (2015). Concurrent movement impairs incidental but not intentional statistical learning. *Cognitive Science, 39*(5), 1081-1098.

Strait, D. L., & Kraus, N. (2011). Playing music for a smarter ear: cognitive, perceptual and neurobiological evidence. *Music Perception: An Interdisciplinary Journal, 29*(2), 133.

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hearing Research, 261*(1), 22-29.

Strouse, A., & Wilson, R. H. (1999). Recognition of one-, two-, and three-pair dichotic digits under free and directed recall. *Journal of the American Academy of Audiology, 10*(10), 557-571.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research, 161*(1), 1-10.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychological Review, 107*(4), 885-913.

Turk-Browne, N. B., Junge, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134*(4), 552-564. doi: 10.1037/0096-3445.134.4.552

Turner, A., & Sandford, J. (1995). *A normative study of IVA: Integrated visual and auditory continuous performance test.* Paper presented at the Annual Convention of the American Psychological Association.

Van der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2006). The Stroop color-word test: influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment, 13*(1), 62-79. doi: 10.1177/1073191105283427

Yost, W. A. (1996). Pitch strength of iterated rippled noise. *The Journal of the Acoustical Society of America, 100*(5), 3329-3335.

Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging, 27*(2), 410. doi: 10.1037/a0024816

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PloS one, 9*(6), e99868. doi: 10.1371/journal.pone.0099868

This page is intentionally left blank

Chapter 3

Musicians' online performance during auditory and visual statistical learning tasks

Pragati Rao **Mandikal Vasuki**[1, 2, 3*], Mridula **Sharma**[1, 2], Ronny **Ibrahim**[1, 2], Joanne **Arciuli**[2, 4]

[1] Department of Linguistics, Macquarie University, Sydney, NSW, Australia
[2] The HEARing CRC, The University of Melbourne, VIC, Australia
[3] ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, NSW, Australia
[4] Faculty of Health Sciences, University of Sydney, Sydney, NSW, Australia

* Corresponding author

Pragati Rao **Mandikal Vasuki**

Department of Linguistics (Audiology),

Australian Hearing Hub,

Macquarie University, Sydney Australia.

pragati.mandikal-vasuki@mq.edu.au

Number of Figures: 5

Running title: Online SL in musicians

Status of the paper: Submitted to Frontiers in Human Neuroscience [Manuscript ID-240666] (under review)

The reference list and the in-text citations are formatted according to the journal specifications.

Abstract

Musicians' brains are considered as a functional model of neuroplasticity due to the structural and functional changes associated with long-term musical training. In this paper, we studied statistical learning, which refers to implicitly extracting statistical regularities from a continuous stream of stimuli. Over the course of two experiments, we investigated if long-term musical training is associated with better extraction of statistical cues in an auditory statistical learning (aSL) task and a visual statistical learning (vSL) task – both using the embedded triplet paradigm. Online measures, characterized by event related potentials, were recorded while participants were exposed to a continuous stream of individually presented pure tones in the aSL task or individually presented cartoon figures in the vSL task. Participants also completed a behavioural forced choice task to assess statistical learning. In Experiment 1, musicians showed advantages when compared to non-musicians in the online measure (early N1 and N400 triplet onset effects) as well as the behavioural measure (better scores) during the aSL task. However, there were no differences between musicians and non-musicians on behavioural and online vSL tasks. Results from the two experiments show that musical training is associated with enhancements in extraction of statistical cues only in the auditory domain.

Keywords: Online segmentation, auditory statistical learning, visual statistical learning, musicians, N400

3.1 Introduction

Long-term musical training has often been associated with positive effects on the encoding of auditory information. For example, musicians have been reported to have larger brain responses to speech (Musacchia et al., 2007), better pre-attentive discrimination of small changes in auditory stimuli (Koelsch et al., 1999), and better skills at organising tones according to changing pitch relations (Van Zuijen et al., 2004). Changes in brain responses can be observed even before they played incorrect keystrokes indicating that musicians have faster detection of error (Maidhof et al., 2009). Musician's brains are often used to examine effects of training indirectly, and are consequently referred to as a model of cortical plasticity (Münte et al., 2002). Hence, studying the differences in brain responses of musicians and non-musicians could assist in understanding of long-term consequences of musical training. One topic of interest is the ability to identify statistical regularities in auditory or visual input, referred to as statistical learning (SL). We investigated whether long-term musical training is associated with enhanced SL by comparing musicians and non-musicians.

SL, a form of implicit learning, is a powerful learning mechanism thought to play a key role in everyday situations such as language processing in children and adults (Kuhl, 2004; Misyak et al., 2010). It was first described by Saffran and colleagues who showed that infants can use statistical regularities such as transitional probabilities to segment continuous sequences of syllables (Saffran et al., 1996a). They used an embedded triplet paradigm for evaluating SL. In this paradigm, participants are first exposed to a continuous sequence of auditory or visual stimuli in which each stimulus is presented one at a time. Unbeknownst to the participants, the continuous sequence is composed into smaller sequences such as triplets. The items within a triplet have a strong statistical probability (or high transitional probability) of co-occurring. Thus, presentation of one item within a triplet strongly predicts presentation of the subsequent item. Boundaries for these ‘embedded’ triplets occur where transitional probability between the two items is low. After being exposed to such a continuous stream for a period of time, referred to as familiarization, participants are assessed on how well they learnt the ‘embedded’ triplets using a behavioural task. Discrimination between embedded and novel triplets is assessed using a habituation paradigm in infants and a forced-choice task in adults. Sensitivity to statistical cues (transitional probabilities) helps in identification of familiar triplets by assisting segmentation of the continuous sequence during the familiarization phase.

In the seminal study by Saffran and colleagues, a stream of syllables was used for familiarization (Saffran et al., 1996a). Since then, SL has been evaluated in different modalities (auditory, visual, and tactile) using a variety of stimuli. Some of the stimuli used to measure auditory statistical learning (aSL) include speech syllables (Saffran et al., 1996b), tones (Saffran et al., 1999), morse code (Shook et al., 2013), and sung language (Schön and François, 2011). Commonly used stimuli to evaluate visual statistical learning (vSL) include geometrical shapes (Fiser and Aslin, 2002), coloured shapes (Kirkham et al., 2002), and cartoon figures (Arciuli and Simpson, 2011; 2012b). Overall, by using various types of stimuli, these studies have demonstrated the robustness of SL mechanism.

An emerging area of research is the association between implicit (statistical) learning and music exposure. Although knowledge of music can be acquired explicitly, it is also acquired implicitly through attending and interacting with a large number of music samples (Rohrmeier and Rebuschat, 2012). A variety of musical structures and features can be learnt implicitly; for instance, timbre sequences (Tillmann and McAdams, 2004), chord sequences (Jonaitis and

Saffran, 2009), and rhythmic patterns (Schultz et al., 2013). The familiarity of these implicitly acquired structures further governs the liking of such structures (Zajonc, 2001). Consequently, SL and wider implicit learning mechanisms may hold the key to learning musical structures and appreciation of music. Better implicit learning primes and sharpens formation of expectancies, and can help in parsing processes that underlie recognition as well as segmentation (Rohrmeier and Rebuschat, 2012). Although it is reasonable to assume that musical training might be associated with better performance on implicit learning tasks, the experimental results do not always concur. Some behavioural studies have reported that musicians were better than non-musicians at learning statistics in a stream of Morse code (Shook et al., 2013), and tone triplets (Mandikal Vasuki et al., 2016). Other studies have shown that musicians and non-musicians had similar performance for learning of unfamiliar music scales (Loui et al., 2010), and learning of a sung language (François and Schön, 2011). Most of the aforementioned studies used a behavioural measure to assess SL. Using a neurophysiological measure along with a behavioural measure can give us a deeper insight into mechanisms of SL in this population. Studying these mechanisms enhances our understanding of experience-driven cortical plasticity in musicians.

SL has been studied using neurophysiological measures such as electroencephalography (EEG), magnetoencephalography (MEG), and near infrared spectroscopy. SL can be studied in two ways using these techniques. Firstly, an online measure of SL can be obtained by recording EEG during the familiarization phase (e.g., Abla et al., 2008). Secondly, EEG can be recorded during the test phase so as to compare the event related potentials (ERPs) obtained in response to familiar and unfamiliar items (e.g., François and Schön, 2011). In particular, the N1-P2 and the N400 regions have been identified as neurophysiological correlates of SL (Cunillera et al., 2006; Abla et al., 2008; Abla and Okanoya, 2009; Cunillera et al., 2009).

A study by Sanders et al. (2002) investigated online measures of SL using continuous speech stream. This stream comprised trisyllabic words (e.g., *babupu, bupada* etc) and was used for familiarization. They showed that word onsets (i.e. initial syllables) elicited larger N1 and N400 potentials. This effect was referred to as the word onset effect. As described previously, a high within-word transitional probability makes the later items within a word more predictable/familiar. In contrast, due to low transitional probabilities at word boundaries, it is difficult to predict the onset of a word (Abla et al., 2008). Thus, larger ERPs are obtained in response to unfamiliar items resulting in a word onset effect. The word onset effect is considered as an evidence of successful segmentation. In a series of studies, Abla and colleagues also recorded ERPs during familiarization phase in adult non-musicians using non-linguistic stimuli. In the first study, an online measure of aSL was obtained as participants listened to three familiarization streams made by concatenation of six pure tone triplets (Abla et al., 2008). In the second study, the tones were replaced by familiar, geometric shapes to assess vSL (Abla and Okanoya, 2009). Based on the performance in the subsequent test phase, participants were divided into high learners, middle learners, and low learners. The high learners showed a triplet onset effect (larger N1 and N400 for the initial stimulus of a triplet) during the first stream in the aSL task. While triplet onset effect was observed in the later streams for the middle learners, it was absent in the low learners for all streams in the aSL task. In the vSL task, however, a triplet onset effect (larger N400) was observed in the first stream for the high learners only.

François and colleagues compared aSL in musicians and non-musicians using ERPs. They recorded ERPs during the test phase of an embedded triplet paradigm using sung language triplets (François and Schön, 2011). They showed that musicians exhibited a familiarity effect (smaller responses for familiar items) at around 200 ms (P2) and at later negativity (around 450 ms) for linguistic and musical stimuli. A subsequent study (François et al., 2014) recorded ERPs during exposure to a familiarization stream of sung language embedded triplets. To explore ongoing brain dynamics as learning takes place, the entire session was divided into 4 time bins and the N400 amplitude across the 4 time bins was compared. Both groups showed an increase in the N400 amplitude in the first time bin. The amplitude increase in both groups was attributed to building up of initial prototypes. The non-musicians showed a linear increase in the N400 amplitude across the rest of the time bins. However, in the musician group, amplitude of the N400 reached an asymptote between the second and third time bin, followed by a decrease in amplitude in the fourth time bin. Thus, an inverted U (increase-asymptote-decrease) learning curve was observed in musicians. The asymptote in the learning curve of musicians was attributed to consolidation of units into templates, while the decrease in amplitude was due to the effect of repetition of templates (familiarity effect). This finding was interpreted as faster segmentation of a sung language stream by musicians.

Interestingly, studies have also demonstrated that musical training may be linked with enhancements in visual processing. For instance, musicians have been reported to have larger grey matter in areas associated with visual processing such as superior parietal cortex (Gaser and Schlaug, 2003). Further, Patston et al. (2007) used latency of N1 responses to measure interhemispheric transfer time (IHTT) in musicians and non-musicians. The IHTT represented the speed of transfer for visual information across the corpus callosum. Non-musicians showed faster IHTT from the right to the left hemisphere than from left-to-right. In contrast, the musicians showed no directional advantage indicating a more balanced visual processing in musicians than in non-musicians. In addition, memory for visual materials, and visual attention is also enhanced in musicians (Rodrigues et al., 2007; Jakobson et al., 2008). However, very little is known about how the learning of regularities in the visual domain (vSL) proceeds in musicians. Whilst previous studies have investigated aSL in musicians using neurophysiological paradigms, there have been few empirical investigations of both aSL and vSL in musicians.

In the present study, we compared online measures of aSL and vSL in musicians and non-musicians by recording ERPs during the familiarization phase. Specifically, over the course of two experiments, we investigated how musicians and non-musicians perform on: a) online segmentation and behavioural tasks assessing aSL; and b) online segmentation and behavioural tasks assessing vSL. To this end, we used an embedded triplet paradigm for assessing unimodal auditory and visual SL. Data for the aSL and vSL tasks was collected in the same session. For ease of description, the aSL and vSL tasks are described as separate experiments. The aSL task and results are described in Experiment 1, while the vSL task and results are described in Experiment 2. All participants performed both aSL and vSL tasks. The order of presentation of aSL and vSL tasks was counterbalanced across participants. Regardless of the order, all participants first completed the familiarization phases for the aSL and vSL tasks. Participants then proceeded to the test phase in the same order as the familiarization. That is to say, a participant who completed vSL familiarization before aSL familiarization, subsequently completed the vSL test phase before the aSL test phase.

In the present study, instead of speech syllables, we used pure tones (aSL) and cartoon figures (vSL) for familiarization. Thus, for the purposes of this study, we refer to the word onset effect as a triplet onset effect consistent with Abla and Okanoya (2009). We hypothesised that we would obtain a triplet onset effect characterized by larger ERP responses (N1 and N400) for the first stimulus compared to the third stimulus within a triplet. In addition, we also performed correlational analyses to examine the associations between online and behavioural measures. ERP effects may be measured through multiple independent analyses of variance (ANOVAs). However, we lose crucial information while using these methods, and the selection process of 'interesting' data or 'grouping electrodes' is open to user biases (Mensen and Khatami, 2013). Moreover, the multivariate nature of electrophysiological data, that is, measuring a physiological signal over a large number of electrodes, at a number of time points, increases family-wise error rates. To overcome these problems, we analysed ERP data using a cluster-based permutation statistical analysis (Maris and Oostenveld, 2007).

3.2 Experiment 1

In the first experiment, we assessed aSL in musicians and non-musicians using pure tone triplets as reported in Saffran et al. (1999). We hypothesized that musicians would show a larger triplet onset effect in the online measure and higher scores on the behavioural task of SL compared to non-musicians.

3.2.1 Materials and methods

3.2.1.1 Participants

Seventeen musicians (mean age of 32 years; SD 13.2) and 18 non-musicians (mean age of 28.9 years; SD 9.3) with normal hearing (defined as $< = 20$ dB HL pure-tone thresholds at octave frequencies from 250 to 8000 Hz), normal to near normal corrected vision, and no history of neurological disorders participated in the study. An independent samples t-test showed that the groups did not differ significantly in age [$t\ (33) = 0.81$ p=0.42, d=0.3].

Musicians were classified as individuals who had learnt music before the age of 9 and had more than 10 years of musical experience. All the musician participants reported that they still actively practised music. Details about the musical and educational background of the participants have been previously described in Chapter 2 (Mandikal Vasuki et al., 2016). Participants who were categorized as non-musicians had minimal to no formal musical training, and did not report playing a musical instrument at the time or routinely participating in any musical activity (other than informal listening). Only 3 non-musicians reported having previous musical experience (less than 3 years, on average).

All participants lived in the greater Sydney metropolitan area, were native speakers of English and were right handed as assessed using Edinburgh Handedness inventory (Oldfield, 1971). The study was approved and conducted under the ethical oversight of the Macquarie University Human Participants Ethics Committee. Written consent was received from all participants. Subsequent to the participation, all participants were provided with a gift voucher towards their travelling expenses.

3.2.1.2 Stimuli and tasks

The SL task was designed based on previously published embedded triplet tasks (Saffran et al., 1996b; Saffran et al., 1999; Arciuli and Simpson, 2011; 2012b) which consisted of a familiarization and a surprise test phase.

3.2.1.2.1 Familiarization phase

The stimuli for aSL familiarization task were created using musical tones from the same chromatic set (beginning at middle C) as previously described (Saffran et al., 1999; Abla et al., 2008). Eleven pure tones were created using MATLAB (R2013a). The tones were 550 ms in duration with 25 ms rise and fall time. Based on the previous study (Abla et al., 2008), we used 550 ms stimulus duration to obtain non-overlapping ERP responses for individual stimuli within a triplet. These tones were combined in succession to form six triplets (ADB, DFE, GG#A, FCF#, D#ED, CC#D). The six triplets were then concatenated pseudo-randomly to form three continuous streams of stimuli (e.g., ADBGG#AD#EDADBFCF#). These three stimuli streams are henceforth referred to as stream 1, stream 2 and stream 3. Following previous studies (Arciuli and Simpson, 2011; 2012b), the triplets were combined with two randomization constraints: a) consecutive repetition of a triplet was not allowed (e.g., ADBADB would not be allowed); and b) consecutive repetition of two triplets in the same order was not allowed (e.g., ADBGG#AADBGG#A would not be allowed). Thus, the stimuli can be regarded as streams where triplets are ‘embedded’. Each stream was made up of 40 repetitions of a triplet. A stream was approximately 7 minutes in length.

There was no significant difference between the frequency of the tones at each position within tone triplets [$F (2, 10) = 1.48$, $p = 0.28$]. The mean frequency of tones within a triplet were as follows: ADB= 409.2 Hz; DFE= 324.2 Hz; GG#A=415.8 Hz; FCF#= 326.9 Hz; D#ED=311.5 Hz; and CC#D=277.5 Hz. There was no significant difference between mean pitch intervals within- versus across-triplets (3.1 vs 4.9 half tones in average). The transitional probabilities (TP) within triplets ranged from 0.25-1 (mean 0.625) whereas the TPs across triplet boundaries were 0.04-0.3 (mean 0.11). At the end of familiarization, a participant would have been exposed to 21 minutes (7 minutes X 3 streams) of aSL stimuli.

EEG data was recorded during the familiarization phase to obtain an online measure of learning. While participants listened to the familiarization stimuli, they also performed a cover task to ensure attentiveness. The cover task was an oddball detection task. The oddball stimulus was a pure tone with a frequency of 1319 Hz. Forty presentations of the oddball stimulus occurred randomly at the end of triplets. To ensure learning was implicit, participants were neither given instructions about the nature of the embedded triplets within the familiarization stream nor told to learn or remember anything. Participants were also unaware of the upcoming test phase.

3.2.1.2.2 EEG recording

EEG and electrooculography (EOG) signals were collected as the participants were exposed to the familiarization streams. Both horizontal (HEOG) and vertical (VEOG) signals were acquired by placing four electrodes: one at the outer canthus of each eye and one below and above the right eye. EEG was recorded using sixty-four electrodes set up according to the international 10-20 system (Jasper, 1958). EEG was recorded using Ag/AgCl sintered electrodes attached to EasyCap® on a Neuroscan system, (Compumedics Inc). The impedance of all electrodes was maintained below 5 kΩ using a combing technique (Mahajan and McArthur, 2010). All data were sampled at 1000 Hz. Triggers were inserted to mark the onset of each stimulus within a triplet.

3.2.1.2.3 Test phase

After the familiarization phase was completed, participants were informed about the surprise test phase. The construction of the 36 trials surprise test was based on previously published research (Saffran et al., 1999). Six novel triplets were created by combining the same previously mentioned eleven pure tones. The constituent tones in novel triplets had never occurred in that order in the familiarization phase. The task was a two-alternative forced choice (AFC) task where each embedded triplet was paired with a novel triplet. The order of presentation of the embedded and novel triplets was counterbalanced. The participants were asked to indicate which of the two triplets was familiar to them through a button press response.

3.2.2 Data analysis

A behavioural index of learning was calculated as the percentage of correctly identified embedded triplets during the test phase. Consistent with previous SL studies (Conway et al., 2010; Arciuli and Simpson, 2012a; Stevens et al., 2015), participants who scored outside the range mean by ± 2 SD were excluded from further analyses. One-sample t-tests were used to determine whether SL performance was significantly different from chance (50%) in each group. We then conducted an independent t-test to compare performance across the two groups (musicians and non-musicians).

ERP analysis was performed only for the participants retained after exclusion based on score deviation. The continuous EEG files were labelled according to the order of presentation – stream 1, stream 2, and stream 3. Ocular artefacts were removed using EOG artefact reduction implemented in Edit module of Neuroscan (Scan 4.5). After ocular artefact removal, EEG was further processed using Fieldtrip toolbox (Oostenveld et al., 2010) implemented in MATLAB (R2014a). The data were re-referenced to the average of the left (M1) and the right (M2) mastoids.

The continuous EEG was divided into 750 ms epochs which ranged from -100 ms to 650 ms relative to the onset of the presented tone. The epochs were then baseline corrected using the mean amplitude of the signal between the -100ms to 0 ms period. Each epoch represented the evoked response to a single stimulus in the embedded triplet. In order to remove noisy trials, a variance rejection criteria was used, trials which had variances of more than $300 \mu V^2$ between

-100ms up to 650 ms were excluded from further analysis. The accepted trials were bandpass filtered with a frequency cut-off between 0.1 to 30 Hz and a transition band roll-off of 12 dB/octave. The filtered trials were averaged to obtain the ERP waveform. To evaluate the triplet onset effect, ERP waveforms for the initial tone (T1) and final tone (T3) of embedded triplets were compared in each group.

Non-parametric randomization procedure was used to overcome the problem of multiple comparisons over a large group of electrodes (Maris, 2004; Maris and Oostenveld, 2007). The mean amplitude (μV) values for each individual stimulus in time bins of 1 ms starting from 40 ms after the onset of a trigger were taken was input. The statistical analysis produced the following outputs: a cluster of electrodes in which the difference between the conditions tested was significant in each time bin; the sum of t statistics in that cluster; and Monte Carlo estimates of *p*-values. This Monte Carlo estimate is obtained by calculating the test statistic a large number of times and comparing these random test statistics (i.e., draws from the permutation distribution) with the observed test statistic. The Monte Carlo estimate of the permutation *p*-value is the proportion of random partitions in which the observed test statistic is larger than the value drawn from the permutation distribution. In all our analyses, the Monte Carlo *p*-values were calculated on 1000 random partitions. The output is considered corrected for multiple comparisons as only those clusters will be identified that have higher cluster values than 95% of all clusters derived by random permutation of data. This process was implemented using the Fieldtrip toolbox and custom MATLAB scripts. Consistent with previous research (Abla et al., 2008), to compare the performance in the online segmentation (aSL) task, we measured the triplet onset effect (T1 vs. T3) for each stream in the two groups.

In order to check if the individual differences in ERP responses were associated with behavioural performance on the SL task, we performed a brain-behaviour correlation for all the participants. As triplet onset effect indicates successful segmentation, we calculated the difference between ERPs evoked by the initial stimulus and final stimulus within a triplet for each stream in the SL task (T1 minus T3). Then, the correlation procedure implemented in BESA statistics 2.0 was applied to test the association between the difference waveform and the behavioural SL score. The inputs for this procedure were the ERP difference waveforms (0 to 650 ms) and the behavioural SL scores. Correction for multiple comparisons over a large group of electrodes was performed using data clustering and permutation testing (Maris and Oostenveld, 2007). This process tested for a relationship between behavioural test phase results and ERPs obtained during the familiarization phase.

3.2.3 Results

3.2.3.1 Behavioural SL

After removal of outliers as described in previous section, a total of 34 participants were retained for aSL (17 musicians). Participants in both groups responded correctly to the oddball stimulus with over 80% accuracy. The mean performance of musicians was 78.3 % (SD 7.8) and non-musicians was 67.2 % (SD 11.5) on the behavioural aSL task. Both groups performed significantly above chance on the aSL task [one sample t-test; musicians $t(16) = 14.9$, $p < 0.001$, $d = 3.62$; non-musicians $t(16) = 6.1$, $p < 0.01$, $d = 1.49$]. An independent samples t-test showed that musicians outperformed non-musicians in the aSL task [$t(32) = 3.29$, $p < 0.01$, $d = 1.13$]. In addition, an item analysis showed that responding was consistent across all the six triplets in both groups.

3.2.3.2 Event related potentials

ERPs were used to obtain an online measure of segmentation capability in musicians and non-musicians. Figure 1 shows the grand-averaged ERP waveforms and topographies elicited in response to initial (T1) and final (T3) tones of a triplet across the three streams for musicians and non-musicians. In both groups, we observed an N1 component peaking at approximately 100 ms, and a P2 component peaking at approximately 200 ms followed by an N400 component between 300-500 ms.

3.2.3.2.1 Musicians

The amplitude of the N1component was significantly larger for T1 than T3 in the first stream ($p<0.005$). The significant clusters were distributed centrally. This difference was not observed in streams 2 and 3 (all contrasts $p>0.05$).

An N400 triplet onset effect, where amplitude of the N400 was larger for T1 than T3, was observed in all three streams (stream 1: $p<0.005$; stream 2: $p<0.01$; stream 3: $p<0.01$). As seen in Figure 1, cluster permutation statistics showed significant clusters (represented by asterisks) between 300-500 ms with a centro-parietal distribution. The effect was reduced in stream 3 with significant clusters found between 420-480 ms.

3.2.3.2.2 Non-musicians

In contrast to the musicians, there was no significant difference in the amplitude of the N1 component for T1 and T3 across the first two streams (all contrasts; $p>0.05$) (Figure 1) in non-musicians.

Testing for an N400 triplet onset effect in the latency range from 350 to 500 ms post-stimulus, the cluster-based permutation test revealed no significant difference between T1 and T3 in stream 1 (p>0.05). Interestingly, a significant difference was observed for this latency range in streams 2 and 3 (stream 2: $p<0.005$; stream 3: $p< 0.05$). This effect was most pronounced over the centro-parietal electrodes.

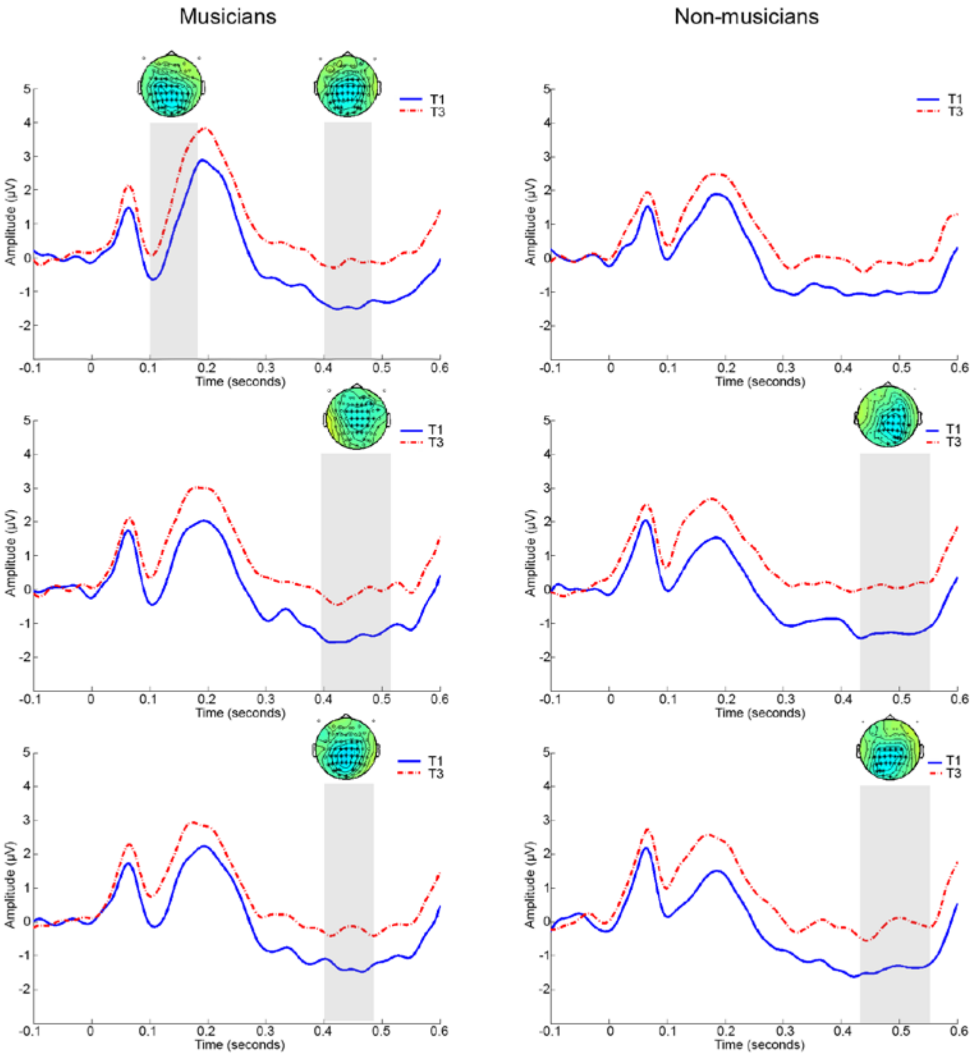

Figure 1: Grand-averaged ERP waveforms for central electrodes sites for musicians (left panels) and non-musicians (right panels) in the aSL online task. The topoplots show significant clusters between initial tone (T1) and final tone (T3) of a triplet. Top panel= stream 1; middle panel= stream 2; bottom panel = stream 3.

3.2.3.3 Brain-behaviour correlation

Correlational analysis revealed a significant association between the difference waveform, that is ERPs for T1 minus T3 and the behavioural SL performance in streams 1 and 2 (stream 1 and 2: $p<0.05$). The significant clusters with a central distribution were found between 100 ms and 400 ms (Figure 2).

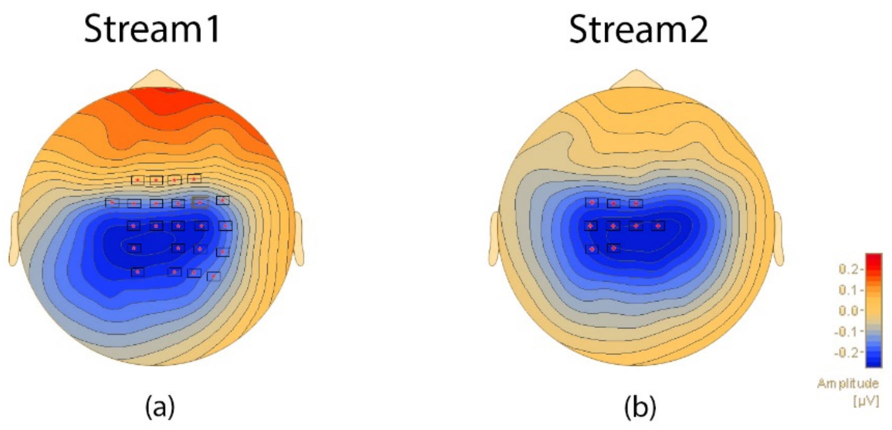

Figure 2: Topoplots showing cluster of electrodes where significant correlations were obtained between behavioural scores and online measures of aSL.

3.2.4 Discussion

At the outset of this experiment, we hypothesized that musicians would show a larger triplet onset effect in the online measure, and score higher on the behavioural task of SL compared to non-musicians. The analysis of data collected during the experiment showed that, when compared with non-musicians, musicians indeed scored higher on the behavioural index, and performed better on being able to segment a continuous tone stream during the ERP task. Better performance on the ERP task was characterized in musicians by presence of both N1 and N400 triplet onset effects during the initial part of familiarization (stream 1). An interesting finding is the appearance of the N400 triplet onset effect only in the later parts of familiarization (streams 2 and 3) for the non-musician group. Consistent with previous literature (Sanders et al., 2002; Sanders et al., 2009), the N400 triplet onset effect was distributed in the centro-parietal regions.

In the current study, musical pure tones were used to construct embedded triplets. In order to recognize that continuous streams were made up of these embedded triplets, participants relied on transitional probabilities. Participants in both groups were able to use these cues and learn the embedded triplets. However, musicians outperformed non-musicians. Better performance of musicians in the behavioural aSL task may indicate that long-term musical training primes musicians to form associations between successive stimuli and thus helps them in making better use of the transitional probabilities. These findings are also supported by a recent longitudinal study which reported that children learning music for 2 years can significantly improve their aSL abilities when compared to a control group (François et al., 2013).

We used ERPs to obtain an online measure of segmentation in this population. When listening to continuous sequences of sounds, elements in the initial part of a sequence elicit a larger negativity. Reduction of both N1 and N400 components for predictable stimuli has been reported to be a ‘marker’ or ‘index’ for online segmentation of continuous sound sequences (Sanders et al., 2002; Sanders et al., 2009). The presence of the N1 and N400 triplet onset effects could arise from one of the two possibilities: (i) differences in processing of initial and final tones of a triplet or (ii) the process of segmentation itself. Findings from Sanders et al. (2009) study show that the N1 and N400 triplet onset effects cannot be solely attributed to acoustic differences between stimuli. Moreover, in our stimuli there was no significant difference between the frequencies of tones at each position within the triplets. The second explanation is supported by research that shows that triplet onset effect is reflects sensitivity to statistical regularities and not acoustic differences between the stimuli (Astheimer and Sanders, 2011).

Interestingly, only the musician group showed an N1 triplet onset effect during the early part of the familiarization sequence (stream 1). Previous studies have shown that the N1 triplet onset effect may be seen in the participants who were classified as ‘expert’ or ‘high’ learners based on performance in the behavioural task (Sanders et al., 2002; Sanders et al., 2009). These findings were attributed to use of additional resources such as selective attention during learning. Additionally, Abla et al. (2008) also reported the presence of an N1 triplet onset effect for their high learner group (i.e. those who performed above mean+0.5 SD on the behavioural task) only in session 1. Taken together, these findings suggest that musicians, when compared with non-musicians, perform as ‘high learners’ or ‘expert listeners’ and possibly use additional strategies such as selective attention for successful recognition of statistical regularities.

Additionally, only the musicians exhibited the N400 triplet onset effect in the first stream. This finding indicates that musicians were able to utilize the statistical structure of the stream and could segment it faster than non-musicians. There are several explanations for these findings. As each item is heard, an increase in exposure presumably helps in formation of the triplet (three-tone) template by computation of transitional probabilities. Working memory resources are helpful during the consolidation of these templates (Cunillera et al., 2009; Lopez-Barroso et al., 2011). Musicians demonstrate faster update of working memory (George and Coch, 2011) and increased neural activity during working memory tasks (Pallesen et al., 2010). In addition, previous studies have shown that musicians are also better at grouping and processing complex auditory patterns (Van Zuijen et al., 2004; Boh et al., 2011). Thus, it is plausible that musicians may be relying more on these abilities than non-musicians to achieve segmentation.

The N400 component is also regarded as an indicator of successful segmentation of the continuous sequence (Cunillera et al., 2009). The appearance of the N400 triplet onset effect in the later streams (streams 2 and 3) in non-musicians may indicate that non-musicians require larger periods of familiarization to successfully segment sequences. This finding is consistent with the ERP effects seen in the middle learner group during the previous study by Abla et al. (2008). Overall, the appearance of N400 triplet onset effect in early streams in musicians and in the later streams in the non-musicians further supports the notion that musicians may be faster at segmentation tasks (i.e. detection of statistical regularities).

3.3 Experiment 2

In Experiment 1, we looked at the neurophysiological mechanisms underlying online segmentation of a continuous tone sequence. In Experiment 2, we explored the mechanisms underlying online segmentation of visual stimuli in the same participants by replacing auditory stimuli with visual stimuli. Subsequently, we obtained a behavioural measure of learning using a forced choice task. Following reports of musicians' enhanced performance on visual tasks (Jakobson et al., 2008; Anaya et al., 2016), we hypothesized that musicians would outperform non-musicians in the behavioural and online vSL tasks.

3.3.1 Materials and methods

3.3.1.1 Participants

All participants from Experiment 1 also completed Experiment 2.

3.3.1.2 Stimuli and tasks

As in Experiment 1, an embedded triplet task was used to assess vSL. The SL task consisted of a familiarization and a surprise test phase.

3.3.1.2.1 Familiarization phase

The vSL task was designed identically to the aSL task in Experiment 1. Eleven cartoon-like figures best described as aliens used by Arciuli and colleagues (see appendix in Arciuli and Simpson, 2011) replaced the 11 pure tones used to construct aSL stimuli. The cartoon figures were rescaled to have equal height and width. These cartoon figures were combined in succession to form six triplets. The six triplets were concatenated pseudo-randomly to form the three familiarization streams henceforth referred to as stream 1, stream 2, and stream 3 respectively. Each stream contained 240 triplets (40 repetitions X 6 triplets) and was 7 minutes long. Figure 3 illustrates the presentation of two triplets during a vSL familiarization stream. The vSL stimuli were delivered using Presentation software (www.neurobs.com) on a CRT monitor placed 1 metre away from the participant. Each cartoon figure was presented for 550 ms against a black background. The vSL familiarization streams had identical statistical structure as the aSL streams. An oddball detection cover task was used where participants were asked to press a button every time they saw a particular cartoon figure. A total of 40 presentations of the oddball stimulus occurred randomly at the end of triplets within each stream. EEG was recorded while participants watched the familiarization streams.

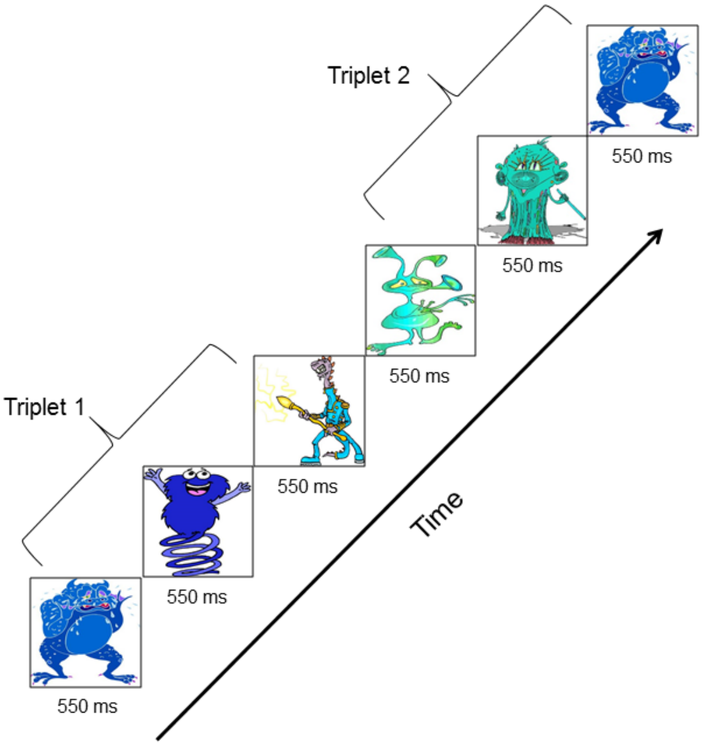

Figure 3: Familiarization stimuli for vSL showing presentation of two cartoon figure 'embedded' triplets.

3.3.1.2.2 EEG recording

The procedure and instrumentation for recording EEG was identical to that described under Experiment 1. Triggers were inserted to mark the onset of each cartoon figure within a triplet.

3.3.1.2.3 Test phase

A 2-AFC test with 36 items was constructed in a similar manner as Experiment 1. Embedded cartoon figure triplets were paired with novel cartoon figure triplets that had never occurred during familiarization. The novel triplets were made up of the same eleven cartoon figures but had never occurred in that order during the familiarization. Participants had to indicate through button press response the triplet that they thought had been presented during familiarization.

3.3.1.3 Data analysis

For the behavioural vSL task, statistical analysis including removal of outliers was performed using the same procedures as described in Experiment 1.

The continuous EEG files were labelled according to the order of presentation – stream 1, stream 2 and stream 3. The ERP analysis followed the same steps as described under Experiment 1. To evaluate the triplet onset effect, ERPs elicited to first picture (P1) and third picture (P3) of a triplet were compared in each stream for musicians and non-musicians. Analysis of vSL surprise test phase results and the brain-behaviour correlation was performed similar to Experiment 1.

3.3.2 Results

3.3.2.1 Behavioural SL

After removal of outliers, a total of 32 participants were retained for analysis (16 musicians). All participants detected the oddball stimulus with an accuracy of above 80%. The mean performance of musicians was 54.3 % (SD 6.5) and non-musicians was 55.5 % (SD 7.6) on the behavioural vSL task. Performance on the vSL task was significantly above chance for both groups [one sample t-test; musicians $t(15) = 2.7$, $p < 0.05$, $d = 0.67$; non-musicians $t(15) = 2.9$, $p < 0.05$, $d = 0.73$]. An independent samples t-test showed that both groups performed similarly on the vSL task [$t(30) = -0.49$, $p > 0.05$, $d = 0.17$].

3.3.2.2 Event related potentials

Figure 4 shows the grand-averaged ERP waveforms and topographies elicited in response to initial (P1) and third (P3) pictures within a triplet across the three streams for musicians and non-musicians. In both groups, a P1 component peaking at approximately 100 ms, an N1 component peaking at approximately 150 ms and a P2 component peaking at approximately 250 ms was seen. The N400 component was not seen in ERPs for either P1 or P3 stimuli. After selecting the a-priori time of interest in our data (N1-P2 region), the cluster-based permutation tests were applied to evaluate the triplet onset effect.

3.3.2.2.1 Musicians

The N1-P2 response to first picture (P1) was significantly larger than the response to the third picture (P3) across all the three streams (all streams: $p<0.001$). Figure 4 shows the significant clusters (represented by asterisks) over the parieto-occipital and occipital electrodes.

3.3.2.2.2 Non-musicians

A significant difference was observed between the P1 and P3 stimuli across all three streams (stream1: $p<0.001$; stream2: $p<0.005$; stream3: $p<0.005$). Similar to the musician group, this effect was observed over the parieto-occipital and occipital electrodes (Figure 4).

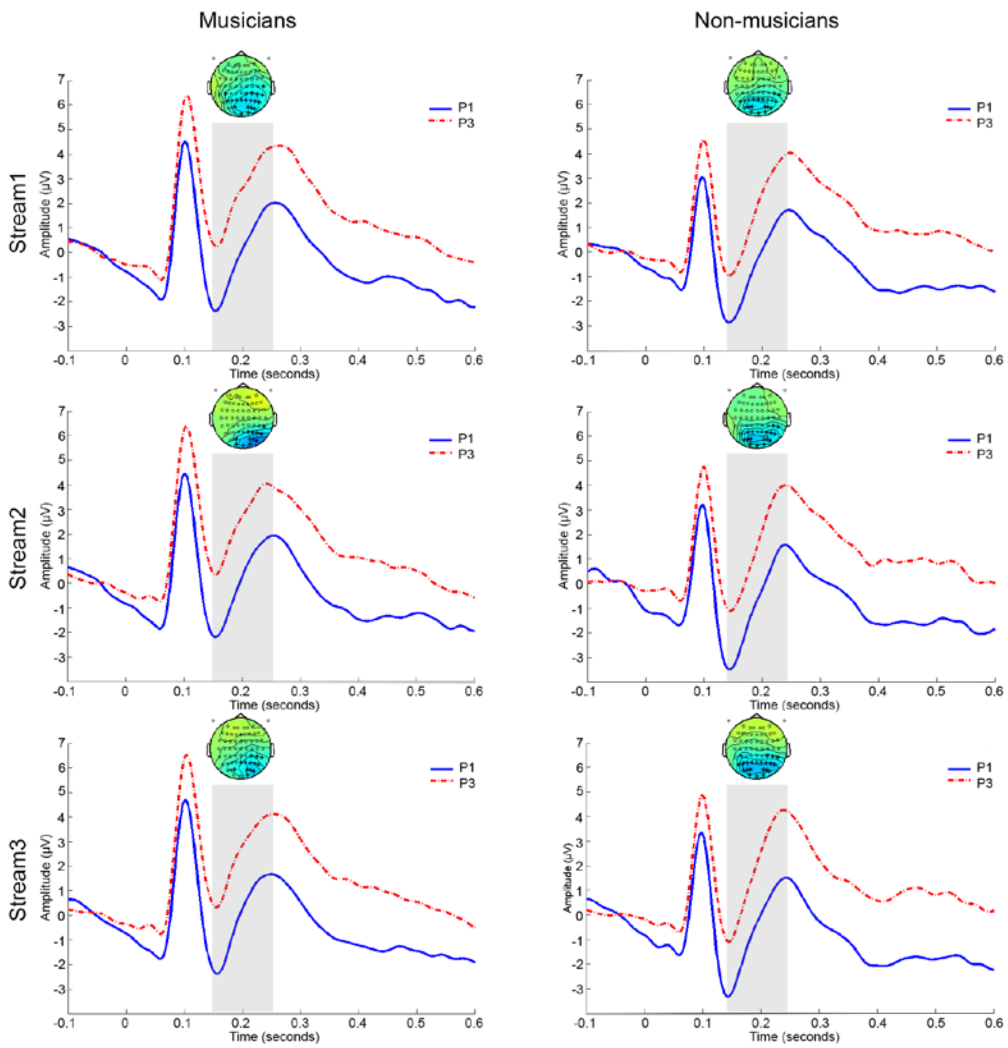

Figure 4: Grand-averaged ERP waveforms for parieto-occipital electrodes sites for musicians (left panels) and non-musicians (right panels). The topoplots show significant clusters between initial picture (P1) and final picture (P3) of a triplet. Top panel= stream 1; middle panel= stream 2; bottom panel = stream 3.

3.3.2.3 Brain-behaviour correlation

Behavioural performance on the vSL task was significantly correlated with the difference waveform obtained by subtracting the ERPs for P1 minus the ERPs for P3 stimulus in streams 1 and 2 only (p <0.05). Significant clusters with a parieto-occipital distribution were found between 200 ms and 550 ms (Figure 5).

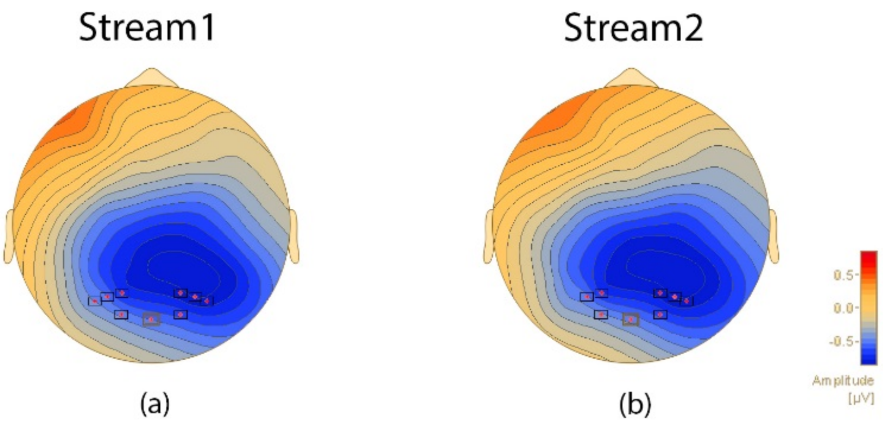

Figure 5: Topoplots showing cluster of electrodes where significant correlations were obtained between behavioural scores and online measure (the difference ERP waveform i.e. P1-P3) for vSL tasks.

3.3.3 Discussion

In this experiment, we explored if musical expertise was associated with better performance on the behavioural and online vSL tasks. We hypothesized that musicians would outperform non-musicians in the behavioural and online vSL tasks. Our hypothesis was rejected as musicians and non-musicians had similar performance on the behavioural and online vSL tasks. The N400 response was not seen in the ERPs to visual stimuli in either group. However, both groups showed a significant reduction in the N1-P2 amplitude as a function of stimulus position, that is, larger N1-P2 responses for the first picture (P1) compared to the third picture (P3) within a triplet.

We evaluated vSL using an embedded triplet paradigm where the individual stimuli were coloured cartoon figures. Learning of these visual 'embedded' triplets was significantly above chance in both groups. It is interesting to note that the mean performance for both groups were lower than those reported in a study using a similar paradigm (Abla and Okanoya, 2009). One explanation for this could be that we used unknown cartoon figures (abstract shapes i.e. aliens) in our vSL instead of familiar, nameable, geometric shapes that were used for vSL in the Abla study. Thus, for the shapes used in our study, participants could not rely on a verbal encoding strategy during learning. For example, if the stimuli can be verbalized, participants may use cues other than statistical cues for learning (for further discussion see Conway and Christiansen, 2005 experiments 1B and 1C). There is some evidence in literature that musicians exhibit advantages in processing of visual stimuli. For example, better design learning (Jakobson et al., 2008) and sequential learning of visuo-spatial patterns (Anaya et al., 2016) has been reported in musicians. However, we did not observe a musicianship advantage in our cohort for learning of visuo-temporal sequences.

While Abla and Okanoya (2009) reported presence of an N400 triplet onset effect between the ERP responses for first and third shape of the shape-words used in their study, we did not observe an N400 triplet onset effect in our dataset. There are two possible explanations for the difference in ERP morphologies between the two studies. Firstly, the appearance of specific ERP components may depend on difficulty level of the stimuli used. Comparison of mean performance of subjects in the behavioural task in the two studies (current study: 54.95%; Abla & Okanaya, 2009: 72.2%) suggests that the vSL task used in the current study may be more difficult. ERP components have been shown to be related to task difficulty and task threshold (e.g., see Caryl and Harper, 1996). In contrast to the aSL task, it is possible that the lack of group difference seen in the vSL task was due task difficulty and consequent lack of sensitivity. An alternative explanation could be related to the use of black and white images (e.g., Abla and Okanoya, 2009) versus coloured images (current study). The current study used complex and coloured images (cartoon figures) for eliciting visual ERPs. ERP responses obtained for rapid presentation of coloured images typically present with a positivity (around 100 ms), followed by a negativity (around 175-200 ms) and another broad positivity (centred around 250 ms) in the posterior electrode sites (e.g., Schupp et al., 2004). A similar pattern was observed in the ERPs in the current study (Figure 4). Overall, the significant brain-behavioural correlations over the parieto-occipital region suggest that, regardless of the group, amplitude changes in the N1-P2 region index successful segmentation of sequentially presented visual stimuli (at least for the stimuli used here).

3.4 General discussion

Over the course of two experiments, we explored online measures of SL using unimodal auditory and visual embedded triplet paradigms. SL performance was also measured using behavioural indices obtained through a 2-AFC triplet recognition tasks. While musicians outperformed non-musicians on the behavioural aSL task, no group differences were seen in the vSL task. A similar pattern of results was seen in the online measures of SL. Musicians or 'high' learners showed both N1 and N400 triplet onset effects in the early part of familiarization streams (stream 1) in the aSL paradigm. An N1 triplet onset effect (but not an N400 triplet onset effect) was seen in both groups for the vSL paradigm. Taken together, these findings suggest differential processing of auditory stimuli in our aSL task in individuals with musical training.

Increased exposure to sounds and auditory training through repeated practice may sharpen musicians' processing of sounds. This could facilitate better and faster segmentation of auditory stimuli where probabilistic cues are key to segmentation. This is reflected as larger differences in the ERP components for initial tones than final tones within a triplet, as well as higher scores on the aSL triplet recognition task. However, an alternative explanation could be that the ERP effects observed in the current study are due to modulation of the level of attention (Daltrozzo and Conway, 2014). Indeed, in a previous study, musicians displayed enhanced top-down modulatory attentional effects in ERPs elicited to auditory stimuli

(Tervaniemi et al., 2009). Moreover, another study showed that musicians recruit more neuronal networks that sustain attention and cognitive control (Pallesen et al., 2010). The enhancements in attention and cognitive control may be used to better integrate information during segmentation tasks compared to non-musicians. These explanations may not necessarily be mutually exclusive but may operate in a complementary manner thereby facilitating segmentation of auditory stimuli in musicians.

Although music is a multimodal learning experience, enhanced SL was only seen in one modality (auditory not visual). Studies in artificial grammar learning also report limited transfer of learning across modalities (Tunney and Altmann, 1999). However, caution must be observed when comparing results across modalities because it is very difficult to control the perceptual saliency of the aSL and vSL tasks. Our vSL task was based on probabilistic sequences in the visuo-temporal domain. Previous studies have shown that the visual sense is more adept at processing spatial cues (Conway and Christiansen, 2005; 2009). A recent study showed that musicians may be more adept at extracting statistical information in visuo-spatial stimuli (Anaya et al., 2016). Perhaps tasks involving online processing (ERPs) of probabilistic visuo-spatial sequences could further explore the neurophysiological mechanisms of vSL in individuals with musical expertise.

In summary, our findings show that musical training is associated with enhanced sensitivity to statistical regularities in auditory stimuli. These enhancements were also observed in the online measure of SL using ERPs. By measuring both neurophysiological and behavioural indices of SL in auditory and visual modalities, our findings add to the growing literature on musical expertise and performance on SL tasks.

3.5 Author contributions

PV, MS, JA designed the study. PV collected the data. PV, RI analysed the data. PV wrote the paper. All authors contributed in revising the paper critically for intellectual content.

3.6 Acknowledgements and Conflict of Interest Statement

This work was funded by the Australia Awards Scholarships and the HEARing CRC, established and supported under the Cooperative Research Centres Program – Business Australia. This work was also supported by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021). Arciuli was supported by a Future Fellowship granted by the Australian Research Council (FT130101570). The authors declare no conflict of interest.

3.7 References

Abla, D., Katahira, K., and Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience* 20(6)**,** 952-964.

Abla, D., and Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research* 64(2)**,** 185-190. doi: http://dx.doi.org/10.1016/j.neures.2009.02.013.

Anaya, E.M., Pisoni, D.B., and Kronenberger, W.G. (2016). Visual-spatial sequence learning and memory in trained musicians. *Psychology of Music***,** 1-17. doi: 10.1177/0305735616638942.

Arciuli, J., and Simpson, I.C. (2011). Statistical learning in typically developing children: the role of age and speed of stimulus presentation. *Developmental Science* 14(3)**,** 464-473. doi: 10.1111/j.1467-7687.2009.00937.x.

Arciuli, J., and Simpson, I.C. (2012a). Statistical learning is lasting and consistent over time. *Neuroscience Letters* 517(2)**,** 133-135. doi: 10.1016/j.neulet.2012.04.045.

Arciuli, J., and Simpson, I.C. (2012b). Statistical learning is related to reading ability in children and adults. *Cognitive Science* 36(2)**,** 286-304. doi: 10.1111/j.1551-6709.2011.01200.x.

Astheimer, L.B., and Sanders, L.D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia* 49(12)**,** 3512-3516.

Boh, B., Herholz, S.C., Lappe, C., and Pantev, C. (2011). Processing of complex auditory patterns in musicians and nonmusicians. *PLoS One* 6(7)**,** e21458.

Caryl, P., and Harper, A. (1996). Event related potentials (ERPs) in elementary cognitive tasks reflect task difficulty and task threshold. *Intelligence* 22(1), 1-22.
Conway, C.M., Bauernschmidt, A., Huang, S.S., and Pisoni, D.B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition* 114(3), 356-371. doi: 10.1016/j.cognition.2009.10.009.
Conway, C.M., and Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(1), 24.
Conway, C.M., and Christiansen, M.H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology* 21(4), 561-580.
Cunillera, T., Càmara, E., Toro, J.M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., et al. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* 48(3), 541-553.
Cunillera, T., Toro, J.M., Sebastián-Gallés, N., and Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain research* 1123(1), 168-178.
Daltrozzo, J., and Conway, C.M. (2014). Neurocognitive mechanisms of statistical-sequential learning: what do event-related potentials tell us? *Frontiers in human neuroscience* 8, 437.
Fiser, J., and Aslin, R.N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(3), 458.
François, C., Chobert, J., Besson, M., and Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex* 23(9), 2038-2043.
François, C., Jaillet, F., Takerkart, S., and Schön, D. (2014). Faster Sound Stream Segmentation in Musicians than in Nonmusicians. *PloS one* 9(7), e101340.
François, C., and Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex* 21(10), 2357-2365. doi: 10.1093/cercor/bhr022.
Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience* 23(27), 9240-9245.
George, E.M., and Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia* 49(5), 1083-1094.
Jakobson, L.S., Lewycky, S.T., Kilgour, A.R., and Stoesz, B.M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception: An Interdisciplinary Journal* 26(1), 41-55.
Jasper, H. (1958). The 10/20 international electrode system. *EEG and Clinical Neurophysiology* 10, 371-375.
Jonaitis, E.M., and Saffran, J.R. (2009). Learning harmony: The role of serial statistics. *Cognitive science* 33(5), 951-968.
Kirkham, N.Z., Slemmer, J.A., and Johnson, S.P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition* 83(2), B35-B42.
Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10(6), 1309-1313.
Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience* 5(11), 831-843.
Lopez-Barroso, D., de Diego-Balaguer, R., Cunillera, T., Camara, E., Münte, T.F., and Rodriguez-Fornells, A. (2011). Language learning under working memory constraints

correlates with microstructural differences in the ventral language pathway. *Cerebral cortex* 21(12)**,** 2742-2750.
Loui, P., Wessel, D.L., and Kam, C.L.H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception: An Interdisciplinary Journal* 27(5)**,** 377-388.
Mahajan, Y., and McArthur, G. (2010). Does combing the scalp reduce scalp electrode impedances? *Journal of neuroscience methods* 188(2)**,** 287-289.
Maidhof, C., Rieger, M., Prinz, W., and Koelsch, S. (2009). Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes. *PLoS One* 4(4)**,** e5032.
Mandikal Vasuki, P.R., Sharma, M., Demuth, K., and Arciuli, J. (2016). Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. *Hearing Research* 342**,** 112-123. doi: http://dx.doi.org/10.1016/j.heares.2016.10.008.
Maris, E. (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology* 41(1)**,** 142-151.
Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods* 164(1)**,** 177-190.
Mensen, A., and Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *Neuroimage* 67**,** 111-118.
Misyak, J.B., Christiansen, M.H., and Tomblin, J.B. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science* 2(1)**,** 138-153.
Münte, T.F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience* 3(6)**,** 473-478.
Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences* 104(40)**,** 15894-15898.
Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9(1)**,** 97-113.
Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2010). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience* 2011.
Pallesen, K.J., Brattico, E., Bailey, C.J., Korvenoja, A., Koivisto, J., Gjedde, A., et al. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PloS one* 5(6)**,** e11120.
Patston, L.L., Kirk, I.J., Rolfe, M.H.S., Corballis, M.C., and Tippett, L.J. (2007). The unusual symmetry of musicians: Musicians have equilateral interhemispheric transfer for visual information. *Neuropsychologia* 45(9)**,** 2059-2065.
Rodrigues, A.C., Guerra, L.B., and Loureiro, M.A. (Year). "Visual attention in musicians and non-musicians: a comparative study", in: *Proceedings of the 3rd International Conference on Interdisciplinary Musicology*).
Rohrmeier, M., and Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science* 4(4)**,** 525-553.
Saffran, J.R., Aslin, R.N., and Newport, E.L. (1996a). Statistical Learning by 8-Month-old infants. *Science* 274(5294)**,** 1926-1928.
Saffran, J.R., Johnson, E.K., Aslin, R.N., and Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70(1)**,** 27-52.
Saffran, J.R., Newport, E.L., and Aslin, R.N. (1996b). Word segmentation: The role of distributional cues. *Journal of memory and language* 35(4)**,** 606-621.

Sanders, L.D., Ameral, V., and Sayles, K. (2009). Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia* 47(4)**,** 1183-1186.
Sanders, L.D., Newport, E.L., and Neville, H.J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature neuroscience* 5(7)**,** 700-703.
Schön, D., and François, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology* 2**,** 167. doi: 10.3389/fpsyg.2011.00167.
Schultz, B.G., Stevens, C.J., Keller, P.E., and Tillmann, B. (2013). The implicit learning of metrical and nonmetrical temporal patterns. *The Quarterly Journal of Experimental Psychology* 66(2)**,** 360-380.
Schupp, H.T., Junghöfer, M., Weike, A.I., and Hamm, A.O. (2004). The selective processing of briefly presented affective pictures: an ERP analysis. *Psychophysiology* 41(3)**,** 441-449.
Shook, A., Marian, V., Bartolotti, J., and Schroeder, S.R. (2013). Musical experience influences statistical learning of a novel language. *The American Journal of Psychology* 126(1)**,** 95.
Stevens, D.J., Arciuli, J., and Anderson, D.I. (2015). Concurrent movement impairs incidental but not intentional statistical learning. *Cognitive science* 39(5)**,** 1081-1098.
Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., and Friederici, A.D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *European Journal of Neuroscience* 30(8)**,** 1636-1642.
Tillmann, B., and McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis) similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(5)**,** 1131.
Tunney, R.J., and Altmann, G. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25(5)**,** 1322-1333.
Van Zuijen, T.L., Sussman, E., Winkler, I., Näätänen, R., and Tervaniemi, M. (2004). Grouping of sequential sounds—an event-related potential study comparing musicians and nonmusicians. *Journal of Cognitive Neuroscience* 16(2)**,** 331-338.
Zajonc, R.B. (2001). Mere exposure: A gateway to the subliminal. *Current directions in psychological science* 10(6)**,** 224-228.

Chapter 4

Investigation of statistical learning and auditory processing in children with music training

Pragati Rao **Mandikal Vasuki**[1, 2, 3*], Mridula **Sharma**[1, 2], Ronny **Ibrahim**[1, 2], Joanne **Arciuli**[2, 4]

[1] Department of Linguistics, Australian Hearing Hub, 16 University Avenue, Macquarie University, New South Wales 2109, Australia

[2] The HEARing CRC, 550 Swanston Street, Audiology, Hearing and Speech Sciences, The University of Melbourne, Victoria 3010, Australia

[3] ARC Centre of Excellence in Cognition and its Disorders, Level 3, Australian Hearing Hub, 16 University Avenue, Macquarie University, New South Wales 2109, Australia

[4] Faculty of Health Sciences, University of Sydney, 75 East St, Lidcombe, 1825, Australia

* Corresponding author
pragati.mandikal-vasuki@mq.edu.au

Status of the paper: Submitted to Clinical Neurophysiology [Manuscript ID- CLINPH-D-16-9937] (under review)

The reference list and the in-text citations are formatted according to the journal specifications.

Abstract

Objective: The question of how musical training can influence auditory and cognitive abilities of children is of considerable interest. In the present study, along with auditory and cognitive abilities, we compared the ability to implicitly learn statistical regularities children with and without music training.

Methods: Statistical learning of regularities embedded in auditory and visual stimuli was measured in musically trained and age-matched untrained children between the ages of 9-11 years. In addition to collecting behavioural measures, we recorded electrophysiological measures to obtain an online measure of segmentation during the two statistical learning tasks.

Results: Musically trained children had better performance on melody discrimination, rhythm discrimination, frequency discrimination, and auditory statistical learning. Furthermore, grand-averaged ERPs showed that word onset (initial stimulus) elicited larger responses in the musically trained children during both auditory and visual online tasks. In addition, children's music skills were associated with performance on auditory and visual statistical learning tasks.

Conclusion and significance: Our data suggests that auditory skills such as rhythm perception might facilitate detection of regularities in children. The observed ERP differences suggest that musical training might be associated with better encoding of both auditory and visual stimuli. These results have potential implications for developing music-based remediation strategies for children with learning impairments.

Keywords

Musical training, statistical learning, musical skills, event related potentials

Highlights

- Musically trained children better at auditory but not visual statistical learning
- Musically trained children also showed larger changes in event related potentials recorded during the statistical learning task
- Individual differences in musical abilities are associated with the capacity for statistical learning

4.1. Introduction

There is growing evidence that musical training is associated with benefits in not only auditory related tasks but also other non-trained tasks. For example, musical training is associated with benefits in cognitive processing tasks such as memory, attention, and intelligence (Ho et al., 2003, Strait et al., 2010, Schellenberg, 2011). Some of these effects can also be observed in children who are learning music. For instance, children with music training have reportedly enhanced skills in a wide variety of tasks such as pitch perception (Magne et al., 2006), non-verbal reasoning (Forgeard et al., 2008), executive functions (Moreno et al., 2011), and language-related tasks (Anvari et al., 2002, Moreno et al., 2009, Tsang et al., 2011). In addition, previous studies have also reported evidence of plasticity in brain structure and functioning in children with musical experience. When compared to a control group, children with music training reportedly have larger auditory evoked potentials (Shahin et al., 2004), larger anterior corpus callosum (Schlaug et al., 2009), as well as larger right precentral gyrus and right primary auditory cortex (Hyde et al., 2009).

Research suggests that there might be evolutionary links between music and language in the brain (Wallin et al., 2001) with considerable overlap between neural structures involved in encoding and perception of music and language (Patel, 2011). Furthermore, both music and language consist of perceptually discrete elements that are combined into structured sequences according to highly complex regularities. It is therefore of interest to investigate if musical experience is associated with processes that facilitate language acquisition and comprehension. One such cognitive process that facilitates language acquisition is implicit learning of statistical regularities – statistical learning (SL). A large body of research in implicit learning has focussed on how knowledge of language and music structures may be acquired implicitly (for reviews see Ettlinger et al., 2011, Rohrmeier et al., 2012). In contrast, there is a relatively small body of work on direct comparison of SL abilities in musically trained and untrained individuals. Furthermore, studying these processes in individuals with developing music skills (e.g. children who are participating in musical training) can help us to understand the potential correlates of music practice in other domains of cognition.

SL is one of the processes thought to play a role in segmentation of continuous speech (word segmentation). Saffran et al. (1996) first reported that individuals were able to segment a continuous stream of speech by identifying statistically coherent units. Cross-sectional research comparing SL in adult musicians and non-musicians has shown mixed results. Some studies have not shown any group differences on behavioural SL tasks (François et al., 2011, Paraskevopoulos et al., 2012) whereas others have shown that adult musicians perform better than non-musicians (Romano Bergstrom et al., 2012, Shook et al., 2013). More recently, Mandikal Vasuki et al. (2016) showed that adult musicians outperformed non-musicians on auditory but not visual statistical learning tasks. Research utilizing electrophysiological measures has shown that musicians were better at learning the statistical structure of a sung language than non-musicians (François et al., 2011, 2014). François et al. (2011) reported that musicians had a larger N1 familiarity effect indicating that they could recognize statistically coherent (familiar) items better than unfamiliar items when compared with non-musicians. Additionally, event-related potentials (ERPs) recorded during the familiarization of a sung language showed that the N400 learning curve saturates earlier in musicians than non-musicians, indicating faster detection of statistical regularities in musicians (François et al., 2014).

To date, only one study has investigated SL in children learning music (François et al., 2013). François et al. (2013) investigated auditory SL (aSL) using a longitudinal approach in 8-year old children. Twenty-four children were pseudo-randomly divided into two groups in which each group received either music or painting training for a period of two years. Two teachers, specifically trained in music or painting (one for each type of training), were recruited for the study. Each group was trained for 45 minutes, twice a week in the first year and once a week in the second year. Stimuli used for assessing SL were 4 tri-syllabic words and each word was sung on a particular melodic contour (e.g. pymiso was sung as B3 E4 F4). Both behavioural and electrophysiological measures were used to assess SL. A two-alternative forced choice (AFC) task was used to assess SL behaviourally. Four part-words were created for the 2 AFC task. Each word (referred to as a familiar word because it was presented during familiarization) was paired with a part-word (referred to as a unfamiliar word because it was not presented during familiarization). While aSL improved over time in children learning music, the performance of the children learning painting remained at chance level on the behavioural test. In addition, the familiarity effect (difference in ERPs for familiar and unfamiliar words in the 450-550 ms latency range) was larger in the music group than the painting group. The familiarity effect was interpreted as evidence that children learning music could better identify statistical regularities. The authors concluded that the enhancements in aSL may be attributed to better rhythm perception in children with music training. To the best of our knowledge, no study has looked at individual differences to assess if higher SL is associated with higher musical abilities such as rhythm or melody discrimination. Furthermore, very little is known about visual statistical learning (vSL) in children with music training, although there are a couple of studies that have shown enhanced visual sequence learning in adult musicians (Romano Bergstrom et al., 2012, Anaya et al., 2016). In the present study, we used a cross-sectional approach to examine the association between SL (both aSL and vSL) and musical training in children. We used both behavioural and online electrophysiological measures to assess SL.

Previous studies have reported online ERP measures of aSL and vSL in non-musician adults (Abla et al., 2008, Abla et al., 2009). In these studies, an embedded triplet task was used to assess SL. The embedded triplet task comprised a familiarization and a test phase.

During familiarization, a stream of concatenated triplets was presented to the participants. Items within a triplet had a higher chance of co-occurrence or higher conditional probability and were therefore more predictable. Abla and colleagues reported a triplet onset effect (in the N1 and N400 response) as an index for online measure of segmentation during familiarization. The ERPs were larger in less predictable positions (initial stimulus of a triplet) than in more predictable positions (final stimulus of a triplet) which was referred to as a triplet onset effect. Individuals who performed better in the test phase of the SL task had larger triplet onset effect.

In the present study, our approach was to examine an array of auditory and cognitive behaviours (including SL) in children who had been enrolled in private music lessons for at least 1.5 years prior to participation. A second group of children, matched in chronological age, socio-economic background and parents' education level, served as a control group for evaluating the effects associated with music instruction. Importantly, we investigated individual differences amongst children to assess if higher SL is associated with higher music skills. In order to understand brain dynamics as the process of learning unfolds, we obtained an online measure of segmentation during the aSL and vSL tasks using ERPs. In the present study, we used similar paradigms and the same principles as Abla et al. (2008), to record and analyse the ERP data. In the current study, we used tones and picture stimuli to evaluate aSL and vSL respectively.

The aim of the present study was to compare musically trained and untrained children on measures of auditory processing, cognitive processing, behavioural and online SL. Based on the aforementioned literature, we hypothesized that children with music training would outperform their peers on the auditory and cognitive processing tasks used in the current study. With regard to the SL tasks, we hypothesized that children with music training will have higher scores on the behavioural SL tasks, and larger triplet onset effects on the online ERP tasks.

4.2. Method

4.2.1. Participants

Fifty children (N=50, 23 girls) were recruited through advertisements on websites and group emails. Among them, 25 were musically trained (mean age = 10.5 years old, SD = 0.9 years; 14 girls) and 25 were not musically trained (mean age = 10.1 years old, SD = 0.9 years; 9 girls). There was no significant difference in age between the two groups [$t(48) = 1.4$, $p = 0.2$, $d = 0.44$]. The musically trained and untrained children are henceforth referred as musicians and non-musicians respectively. Musicians had an average of 3.9 years of private musical training. Some children played more than one instrument and their musical background is detailed in Table 1. Non-musicians had not received any formal musical training. All children were right handed as assessed by Edinburgh Handedness Inventory (Oldfield, 1971). All children were native speakers of English and had normal or corrected-to-normal vision, no known language, reading, or cognitive impairments. Normal hearing was confirmed through peripheral hearing screening involving pure tone audiometry, immittance evaluation (to assess middle ear functioning), and the presence of otoacoustic emissions (to assess functioning of outer hair cells) in both ears. All children lived in the greater Sydney metropolitan area. Participants and their legal guardians provided informed consent according to Macquarie University’s Human Ethics Committee guidelines.

Questionnaires were used to obtain demographic, parental education, and parental income information. Mann-Whitney U tests showed that both groups had similar socio-economic background ($U = 232.5$, $p = 0.1$), maternal education ($U = 280.5$, $p = 0.5$), and paternal education ($U = 256.5$, $p = 0.3$).

Table 1. Details about musical background of the children in the musician group

Instrument (s)	**Age of onset (years)**	**Duration of practice (years)**
Flute	8.5	2
Piano, trumpet	6	4.6
Piano, french horn, trumpet, other brass instruments	4.5	5.5
Keyboard, guitar	6.5	5.3
Piano, clarinet	6.8	4.5
Piano, clarinet	7.8	2.5
Flute	8.3	2.4
Violin, drums	7	2.2
Piano, percussion	5.5	5.4
Cello, saxophone	8	3.8
Violin, clarinet, recorder, piano, school choir	6	3.6
Piano, vocals	9	2.1
Piano, percussion, vocals	4	5
Trombone	7.5	1.6
Violin, piano, guitar	2	7
Piano, clarinet	5	4.8
Saxophone	8	1.6
Piano, vocals	5.3	6.5
Piano	7.8	1.6
Piano	6.8	4.4
Drums	7	3.6
Violin, flute	9	2.5
Vocals, piano	8	2.4
Vocals, clarinet, bass clarinet, Guitar, piano	5	6.4
Piano, vocals, recorder	5	6.5
Mean	**6.6 (SD=1.7)**	**3.9 (SD=1.7)**

4.2.2. Tasks

4.2.2.1. Auditory Processing tasks

4.2.2.1.1. Musical skills.

Children completed both the melody and the rhythm subtests of the Musical Ear test (Wallentin et al., 2010). Each subtest had two example trials and 52 experimental trials. On each trial, children heard two sequences, half of which were the same, and half of which differed in pitch on one or more notes (on the melody subtest) or in rhythm (on the rhythm subtest). Children indicated verbally whether the two sequences sounded the same or different. The researcher recorded their responses on a scoring sheet. A percentage correct score was calculated for each subtest. An overall percentage score for music skills was obtained by combining the scores obtained from the two subtests.

4.2.2.1.2. Frequency discrimination at 1 kHz.

The test consisted of three-alternative forced choice (AFC) trials in which 1 kHz pure tone served as the standard stimulus (described in Peter et al., 2014). The variable (target) stimuli were generated with frequencies ranging from 1001 Hz to 1050 Hz in steps of 1 Hz. Stimuli were presented binaurally through headphones at 50 dB HL. Children were asked to identify the interval containing a different signal through a button press response. Thresholds were estimated based on a 2-down 1-up tracking method, estimating the 70.7% correct point on psychometric function (Levitt, 1971). The arithmetic mean of the last three reversals in a block of six was taken as threshold. Log transformation was applied to the thresholds.

4.2.2.1.3. Dichotic digits (3 pairs test).

The pre-recorded dichotic digits test from the Department of Veterans' Affairs (VA) compact disc (VA-CD) was used. Three digits were presented to each ear in succession (total 6). Children were asked to recall digits presented to both ears in any order (free recall condition). The percentage of correct responses obtained in each ear was recorded. The right ear advantage was calculated by subtracting left ear scores from right ear scores.

4.2.2.2. Cognitive Processing tasks

4.2.2.2.1. Memory.

Forward and backward digit span subtests from the Clinical Evaluation of Language Fundamentals, CELF-IV (Australian) (Semel et al., 2006) were administered. Digits were presented through headphones at the rate of one digit per second and at a level of 50 dB HL. Participants were asked to recall the digits in same order (forward span) or reverse order (backward span). Age referenced scaled scores were obtained from raw scores.

4.2.2.2.2. Screening non-verbal intelligence test.

Children completed the Test of Non-verbal Intelligence (TONI-4; Brown et al., 2010). In this test, children were required to complete a pattern using multiple-choice options. The patterns increase in complexity as the test progresses. Raw scores were converted into age-referenced standardized scores using normative data available from the test manual.

4.2.2.2.3. Sustained Attention.

Sustained attention was assessed using the integrated visual and auditory continuous performance task (IVA-CPT; Turner et al., 1995). This is an age- and gender-normed testing tool to diagnose attention deficits in children. Over a span of thirteen minutes, children were asked to click the mouse every time they saw or heard ‘1’. The number ‘2’ was to be ignored. The normed quotient scores are automatically generated by the reporting component of the software.

4.2.2.3. Statistical learning

The embedded triplet tasks used to evaluate aSL and vSL were constructed based on previous literature (aSL: Saffran et al., 1999, Abla et al., 2008) (vSL: Arciuli et al., 2011). SL tasks consisted of a familiarization phase and a test phase. The two tasks were constructed with identical statistical structure where tones were replaced by pictures for the vSL task.

4.2.2.3.1. Familiarization phase.

The aSL stimuli comprised 11 pure tones within the same octave (starting at middle C within a chromatic set). The vSL stimuli comprised of 11 cartoons figures or aliens which were equal in height and width (see Figure 1 in Mandikal Vasuki et al., 2016). Irrespective of the modality, the familiarization sequence was created as follows. Three individual items were combined in succession to form a triplet. Six such triplets were created (ADB, DFE, GG#A, FCF#, D#ED, CC#D). Forty repetitions of each triplet were concatenated pseudo-randomly to form a continuous stream of stimuli with no gaps between the triplets. The triplets were concatenated with two randomization constraints as described in previous studies (Arciuli et al., 2011, 2012b). Thus, a familiarization stream consisted of 240 triplets and was about 7 minutes long. There were no acoustic discontinuities between two triplets in a stream. Sixty-four channel electroencephalogram (EEG) was recorded during familiarization to obtain an online measure of learning. Triggers were inserted at the onset of each stimulus within a triplet for EEG data analysis.

Transitional probabilities (TPs) were the only indicator of triplet boundaries. The within-triplet and across-triplet boundary TPs were calculated using methods published previously (Saffran et al., 1996). The across-triplet boundary TPs ranged from 0.04-0.23 (mean 0.1) and the within-triplet TPs ranged from 0.25 to 1 (mean 0.63).

In order to encourage participants to pay attention to the familiarization streams, a cover task was employed. The cover task was an oddball detection task. For the aSL familiarization stream, a pure tone (1319 Hz) served as the oddball stimulus. In the vSL task, children were asked to press the button if they saw a particular alien figure. The oddball stimulus was randomly presented forty times at the triplet boundaries in each stream. The children were not given any explicit instructions about the nature of the statistical regularities in the aSL and vSL stimuli and were not told to learn or remember anything.

4.2.2.3.2. Test phase.

Our behavioural measure of SL was obtained using a 2 AFC task. This task consisted of 36 trials. In each trial, an embedded triplet was paired with a novel triplet. A total of six novel triplets were created for this task. A novel triplet comprised same individual stimuli (tones or cartoon figures) but in an order that had not occurred during familiarization. Participants were asked to indicate which of the two triplets were familiar via a button press response. A number of practice trials were performed to ensure that children waited for presentation of both embedded and novel triplets before responding. The stimuli chosen for the practice trials never occurred during the familiarization or test phase. The percentage of correctly identified embedded triplets was recorded for each SL task. The test phase for each of the SL tasks lasted around 5 minutes.

4.2.3. Procedure

Data were collected as part of a larger test-battery in two sessions. In the first session, children completed the behavioural auditory and cognitive tasks in an audiometric sound-treated booth. Stimuli were presented through a PC connected to a clinical audiometer. All stimuli were presented at 50 dB HL through headphones. The aSL and vSL tasks were administered in a counterbalanced order in the second session. All children first completed familiarization for both aSL and vSL tasks during which EEG was recorded. Children were informed about the test phase only after completion of both aSL and vSL familiarization phases to ensure that learning was implicit.

During the second session, children sat in a comfortable chair in a Faraday shielded room, 1 metre away from a computer screen. EEG was recorded using sixty-four electrodes set up according to the international 10-20 system (Jasper, 1958). Both horizontal and vertical electrooculogram (EOG) signals were collected by placing four electrodes: one at the outer canthus of each eye and one below and above the right eye. EEG was recorded using

Ag/AgCl sintered electrodes attached to EasyCap®. The impedance of all electrodes was maintained below 5 kΩ using combing techniques described by Mahajan et al. (2010). The EEG was recorded using Neuroscan system (Compumedics Inc) at 1000 Hz sampling rate.

4.2.4. Data Analysis

Behavioural data analyses were performed using IBM® SPSS software. Separate Multivariate Analysis of Variance (MANOVA) were conducted to compare the performance of musicians and non-musicians on auditory and cognitive tasks. All multiple comparisons were corrected using a false discovery rate technique (Benjamini et al., 1995). One-sample t-tests were used to determine whether performance on each SL task was significantly different from the chance level of 50%. We then conducted an independent samples t-test to compare performance between the two groups (musicians and non-musicians) for each SL task. Pearson's correlations were used to investigate the associations between music skills and the SL tasks.

EEG data were processed using Fieldtrip toolbox (Oostenveld et al., 2010) implemented in MATLAB (R2014b) using custom scripts. The data were re-referenced to the average of the left (M1) and the right (M2) mastoids. Ocular artefact correction in the EEG signals was performed using Independent Component analysis (ICA) which utilises a statistical blind source separation approach (Jung et al., 2000). The function of ICA was to identify components that presented maximal temporal statistical independency. This appeared, a priori, as a valid approach to separate neuronal EEG and ocular artefacts because these signals are generated by different uncorrelated processes (Grouiller et al., 2007).

After removal of ocular artefacts, the continuous EEG was divided into trials/epochs of 750 ms with 100 ms baseline. Therefore, each epoch represented the evoked response to a single stimulus in the embedded triplet. The epochs were baseline corrected between -100 ms to 0 ms. In order to remove noisy trials, a variance rejection criteria was used. Trials which had variances of more than 300μV^2 between -100 ms up to 650 ms, were excluded from further analysis. The accepted trials were bandpass filtered with a frequency cut-off between 0.1 to 30 Hz and a transition band roll-off of 12 dB/octave. The filtered trials were averaged to obtain the event-related potentials (ERP) waveform.

Consistent with previous SL research (Abla et al., 2008, Abla et al., 2009), to compare the performance in the online ERP tasks, we measured the triplet onset effect for the aSL and vSL tasks in each group. The initial and final tone ERP waveforms from the aSL paradigm were labelled as TI and T3 respectively. Similarly, ERP responses for the initial and final pictures of a triplet were labelled P1 and P3 respectively. To overcome the problem of multiple comparisons over a large group of electrodes, we used a non-parametric randomization procedure (Maris, 2004, Maris et al., 2007). This analysis technique was adopted for comparing performance between: a) T1 versus T3 stimuli (the triplet onset effect for aSL); and b) P1 versus P3 stimuli (the triplet onset effect for vSL) for each group. The mean amplitude (μV) values for each individual stimulus, in time bins of 50 ms starting from the onset of a trigger, were taken was input. The output from statistical analysis consisted of a cluster of electrodes, the sum of t-statistics in that cluster, and Monte Carlo estimates of *p*-values. The identified cluster of electrodes were corrected for multiple comparisons as only those clusters were selected which had cluster values higher than 95% among all clusters derived by random permutation of data.

4.3. Results

4.3.1. Auditory processing

Comparing the performance of musicians and non-musicians on measures of auditory processing using Pillai’s trace, a significant effect of group was observed [V = 0.5, F (6, 43) =

8.5, $p < 0.001$]. After correcting for multiple comparisons, separate univariate analysis of variance (ANOVAs) on the dependent variables revealed a significant effect of group on melody discrimination [F (1, 48) = 10.1, p=0.003, $\eta_p^2 = 0.17$], rhythm discrimination [F (1, 48) = 7.8, p=0.007, $\eta_p^2 = 0.14$], overall music skill score [F (1, 48) = 15.4, p<0.001, η_p^2 = 0.24], and frequency discrimination for 1 kHz tones [F (1, 48) = 37.8, p<0.001, $\eta_p^2 = 0.44$]. However, non-significant group effects were observed on performance for dichotic digits test (all ps>0.05). To account for the variation in non-verbal IQ, a multivariate analysis of covariance (MANCOVA) was performed with melody discrimination, rhythm discrimination, music skills as dependent variables, non-verbal IQ as a covariate, and group as a between subjects factor. There was a significant effect of group on these variables even when non-verbal IQ was controlled [V = 0.5, F (3, 45) = 13.4, p< 0.001]. The means, standard deviations for performance on tests of auditory processing are listed in Table 2.

Table 2: Mean, SD for performance on tests of auditory processing. Significant results are in bold font (*p*<0.01).

Test	Musicians		Non-musicians	
	Mean	SD	Mean	SD
Melody discrimination (%)	**67.4**	**8.1**	**60.6**	**6.9**
Rhythm discrimination (%)	**66.2**	**7.5**	**60.4**	**7.0**
Music score (%)	**66.8**	**5.4**	**60.5**	**5.9**
Frequency discrimination test (log)	**0.9**	**0.3**	**1.5**	**0.3**
Dichotic digits test (3 pairs) – right ear scores (%)	85.8	7.1	84.9	9.2
Dichotic digits test (3 pairs) – left ear scores (%)	78.8	11.6	74.9	10.9
Dichotic digits test (right ear advantage)	6.9	9.9	8.6	8.7

4.3.2. Cognitive Processing

Using Pillai's trace, there was no significant effect of group on the measures of cognitive processing [V = 0.1, F (5, 44) = 1.4, *p*>0.05]. After correcting for multiple comparisons, the separate univariate ANOVAs on measures of cognitive processing did not show an effect of group. The mean, standard deviations, and statistics for measures of cognitive processing are given in Table 3.

Table 3: Means and SDs for performance on tests of cognition.

Test	**Musicians**		**Non-musicians**		**Statistics**		
	Mean	**SD**	**Mean**	**SD**	**$F_{(1,48)}$**	***p***	**η_p^2**
Forwards digit span (age referenced scaled score)	8.9	2.6	9.6	2.0	0.9	0.36	0.02
Backwards digit span (age referenced scaled score)	11.6	2.2	11.5	2.3	0.06	0.8	0.001
Non-verbal IQ	116.5	9.4	114.0	8.4	1.0	0.3	0.02
Sustained auditory attention quotient	99.3	17.5	89.3	21.7	3.2	0.08	0.06
Sustained visual attention quotient	102.4	20.6	90.2	14.9	5.7	0.02	0.1

4.3.3. Statistical learning

4.3.3.1. aSL

Following data analyses techniques described in previous SL studies (Conway et al., 2010, Arciuli et al., 2012a, Stevens et al., 2015), participants who scored outside the range mean by ± 2 SD were excluded from further SL data analyses. One musician and one non-musician were removed for excessively high scores on the aSL task. Thus, a total of 48 participants were retained for aSL (24 musicians). Figure 1 shows the performance of the two groups on the aSL task.

All participants scored above 85% on the cover task. The mean performance of musicians was 68.9 % (SD 11.5) and non-musicians was 54.7 % (SD 6.9) on the behavioural aSL task. A one sample t-test showed that both groups performed significantly above chance on the aSL task [musicians t (23) = 8.1, *p*< 0.001, d= 1.66; non-musicians t (23) = 6.1, *p*<0.001, d =1.11]. An independent samples t-test showed that musicians outperformed non-musicians in the aSL task [t (46) = 3.31, *p*<0.005, d=0.96]. Additionally, an item analysis showed that responding was consistent across all the six triplets in both groups [musicians F (5, 115) =1.82, *p*>0.05; non-musicians F (5, 115) =2.37, *p*<0.05, all pairwise comparisons *p*>0.05]. Musicians' performance on the aSL task was neither associated with age of onset of musical training (r=-0.2, *p*>0.05) nor the duration of practice (r=0.36, *p*=0.08).

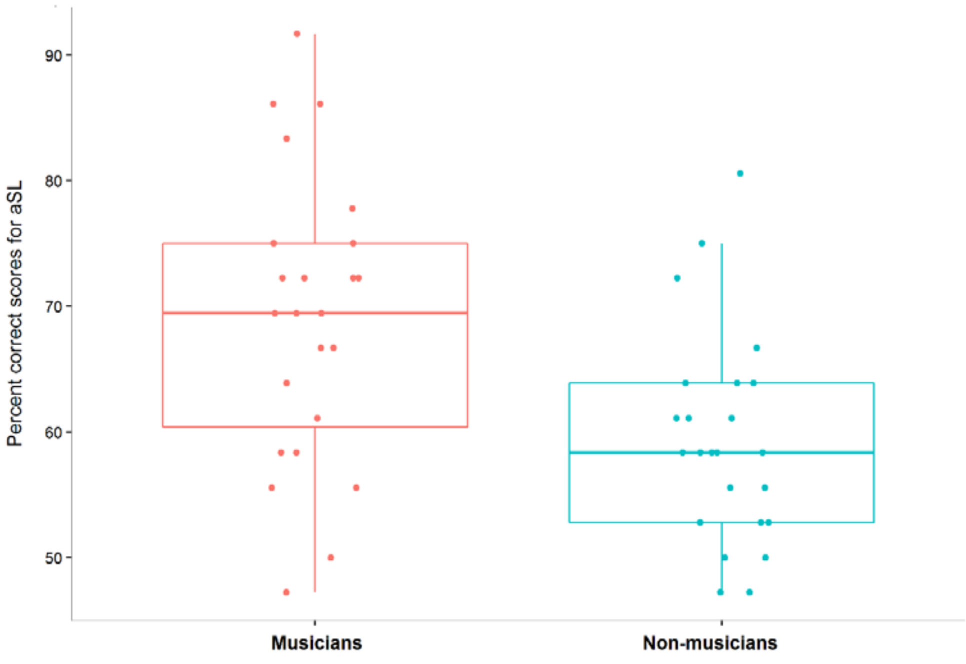

Figure 1: Box-whisker plot showing performance of the two groups on the aSL task. Y axis represents the percentage of correctly identified embedded triplets.

Figure 2 shows the grand-averaged ERP waveforms and topographies elicited in response to initial (T1) and final (T3) tones of a triplet for musicians and non-musicians during the aSL task. In both groups, a P100 component peaking at 100 ms and a N250 component peaking at 250 ms was observed. To measure online segmentation, cluster permutation statistics were applied on the initial (T1) and final (T3) tones of a triplet in each group. For the musician group only, a significantly larger N250 was seen in response to T1 than in response to T3 stimulus (*p*<0.05). This effect was observed in the centro-parietal electrodes.

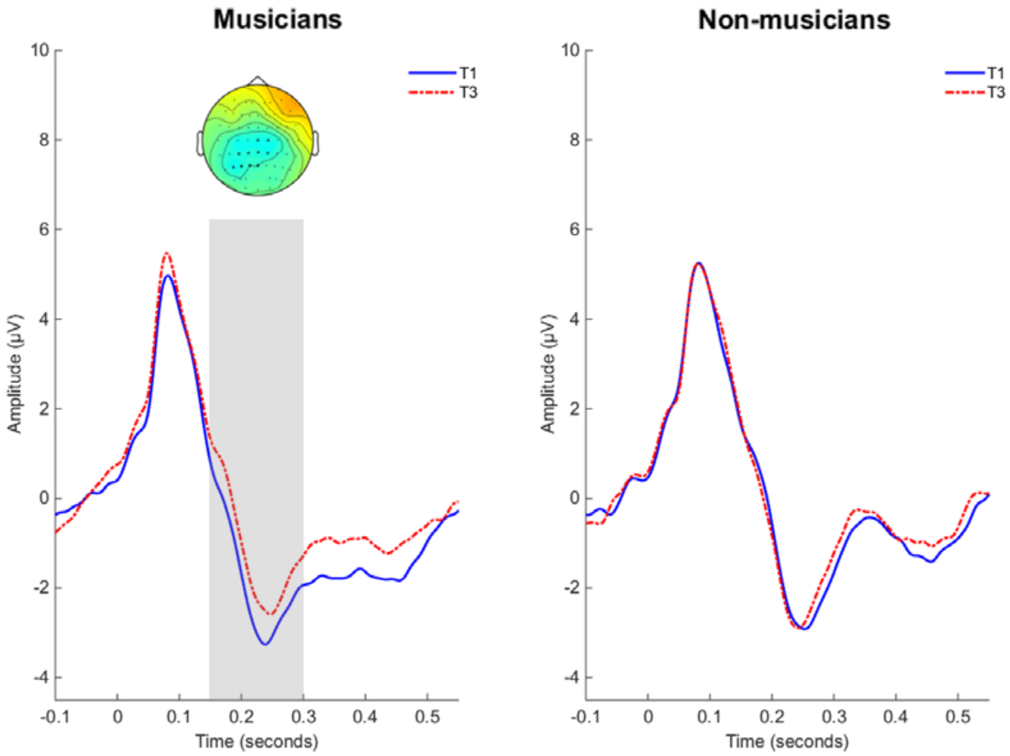

Figure 2: Grand-averaged ERP waveforms for the aSL task at the central electrodes sites. The left and right panels show ERPs for musicians and non-musicians respectively. The shading shows the significant time regions and topoplots show significant clusters between initial tone (T1) and final tone (T3) of a triplet.

4.3.3.2. vSL

Two musicians and one non-musician were removed for excessively high scores on the vSL task. After removal of outliers as described previously, a total of 47 participants were retained for vSL (23 musicians). Figure 3 shows the performance of the two groups on the vSL task. All participants scored above 85% on the cover task. The mean performance of musicians was 54.7 % (SD 6.9) and non-musicians was 53.4 % (SD 5.0) on the behavioural vSL task. A one sample t-test showed performance was above chance in both groups [musicians t (22) = 3.2, *p*< 0.005, d= 0.68; non-musicians t (23) = 3.25, *p*<0.005, d =0.67]. An item analysis showed that responding was consistent across all the six triplets in both groups [musicians F (5, 115) =2.61, *p*<0.05, all pairwise comparisons *p*>0.05; non-musicians F (5, 115) =1.5, *p*>0.05]. There was no significant effect of group on the vSL scores as revealed by an independent samples t-test [t (45) = 0.76, *p*>0.05, d=0.22].

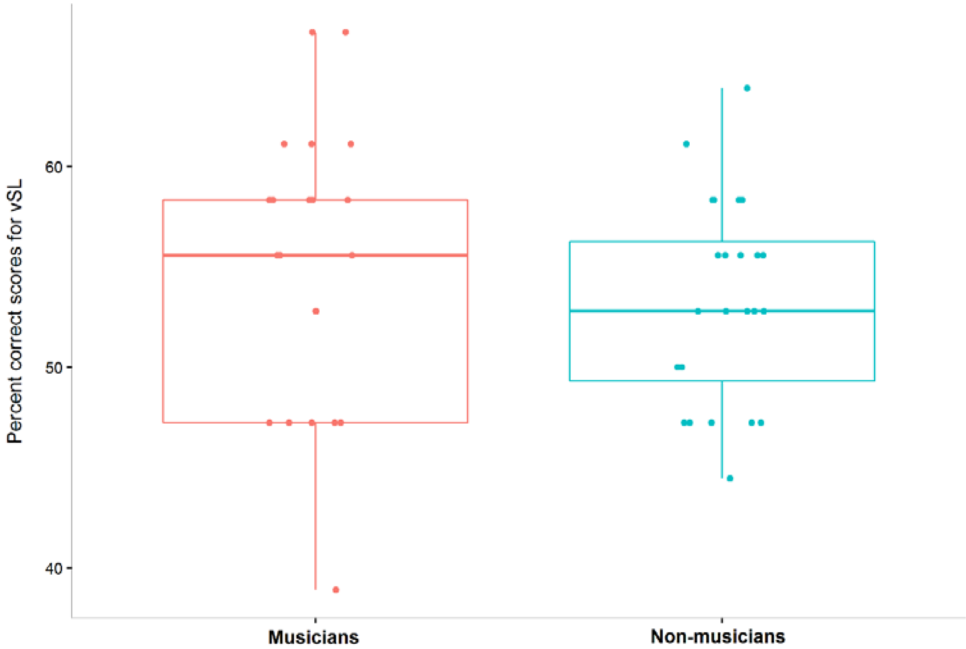

Figure 3: Box-whisker plot showing performance of the two groups on the vSL task. Y axis represents the percentage of correctly identified embedded triplets.

The analyses of ERPs collected during the vSL online measure followed the same procedure as the analyses of the aSL online measure. The morphology of ERPs obtained in response to individual pictures was different than what was obtained in response to tones. Figure 4 shows the grand-averaged ERP waveforms and topographies elicited in response to initial (P1) and final (P3) pictures of a triplet for musicians and non-musicians. In both groups, the first positive peak (P100) was seen at 100 ms, followed by a negative peak (N200) at 200 ms and another positive peak (P300)[1] peaking at 300 ms. Using cluster permutation statistics to measure difference in the ERP responses to initial picture (P1) and final picture (P3) within a triplet, a significant difference was seen in the musician group over the parieto-occipital electrodes in the N200-P300 time region (p<0.005). Although differences in ERP responses to P1 and P3 were observed in the non-musician group, this effect was not statistically significant (p=0.07).

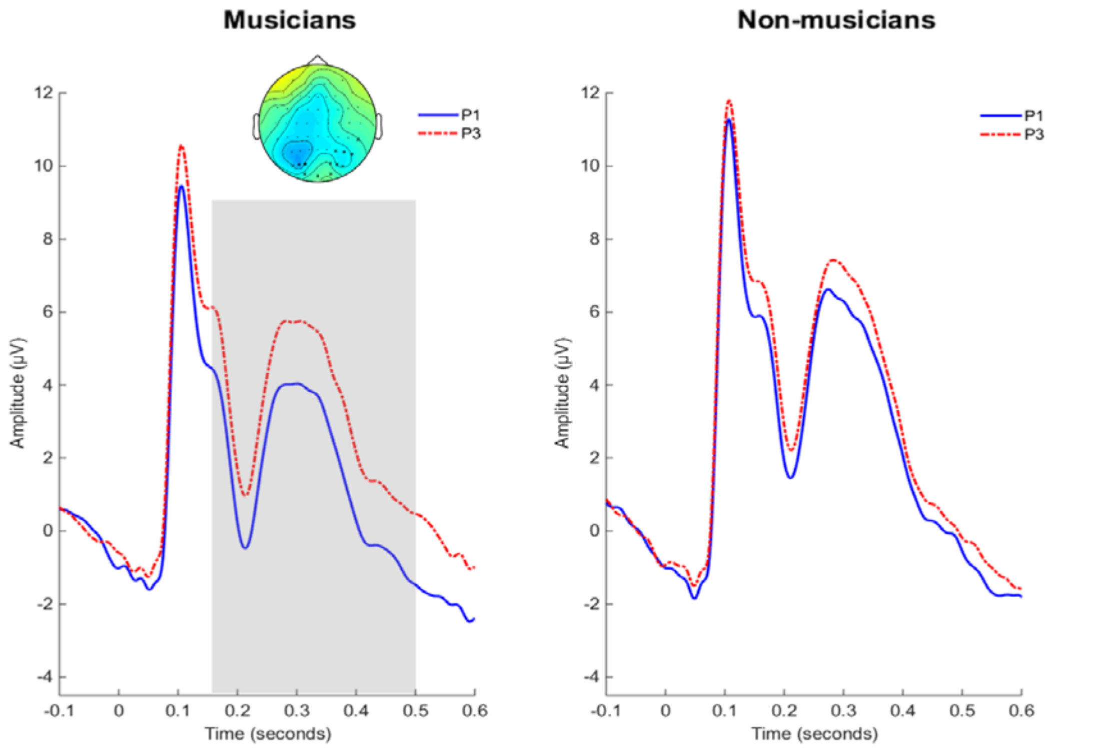

Figure 4: Grand-averaged ERP waveforms for the vSL task at the parieto-occipital electrodes sites. The left and right panels show ERPs for musicians and non-musicians respectively. The shading shows significant time regions and the topoplots show significant clusters between initial picture (P1) and final picture (P3) of a triplet.

[1] It should be noted that peaks are labelled according to the obtained latency and P300 here should not be confused with the P300 peak seen for attention/discrimination paradigms (for e.g. Polich, 1986).

4.3.4. Music skills and behavioural SL

We performed an exploratory correlational analysis in order to assess the association between the scores on behavioural SL tasks and music skills. Scores on melody discrimination and rhythm discrimination tasks were very strongly correlated with the overall music score (melody discrimination: r=0.8, p<0.001; rhythm discrimination: r=0.8, p<0.001). To increase power in the statistical analysis, we computed Pearson's correlations using overall music scores and scores on the two SL tasks for all participants. A moderate positive correlation was observed between aSL and music scores (r=0.41, p<0.005).[2] A weak correlation was also observed between vSL and music scores (r=0.31, p<0.05). Figure 5 shows a scatterplot of performance on the two SL tasks and music scores.

[2] No significant correlation was observed between aSL and overall music scores for either group when group-wise correlations were performed [musicians: r=0.29, p>0.5; non-musicians: r=0.23, p>0.5].

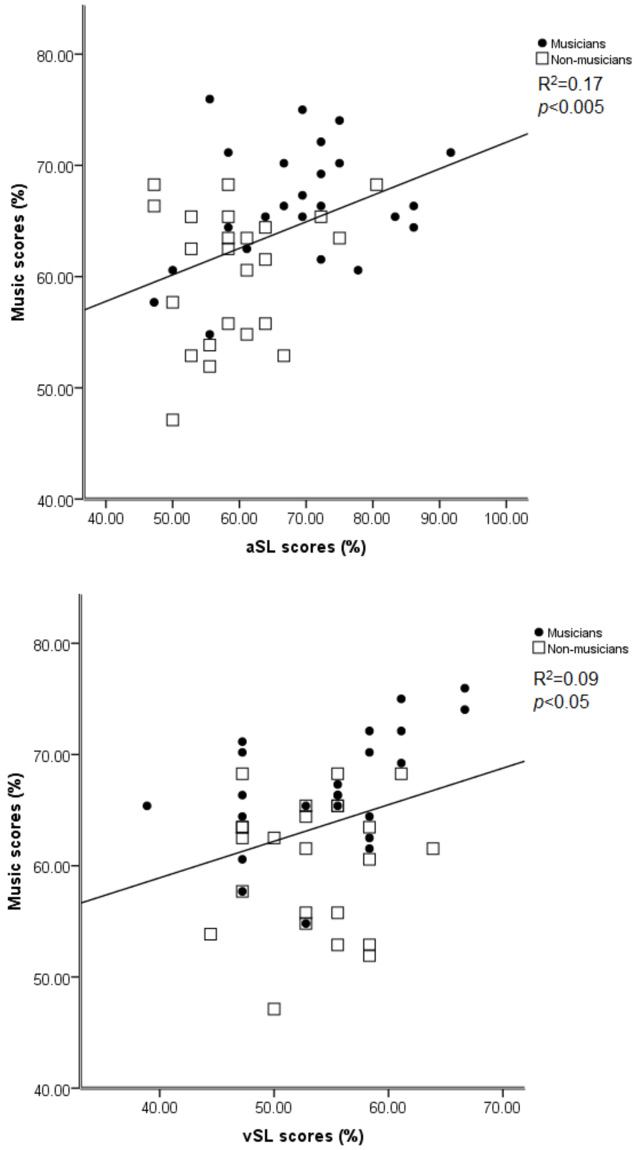

Figure 5: Scatterplot showing percentage of correct triplets identified in the SL task on the X axis and overall scores for the music task on the Y axis. Top panel=aSL task and bottom panel= vSL task. Filled circles represent musicians and unfilled squares represent non-musicians.

4.4. Discussion

The present study compared child musicians and non-musicians on an array of auditory processing, cognitive processing and SL measures. We also obtained behavioural and online measures of aSL and vSL in this population. We hypothesized that musicians would outperform non-musicians on auditory and cognitive processing measures. We also hypothesized that musicians would score higher on the behavioural SL tasks and show larger triplet onset effects in the online ERP tasks.

Our findings confirmed some of our hypotheses; musicians indeed outperformed non-musicians on auditory processing tasks such as melody discrimination, rhythm discrimination, and frequency discrimination tasks. In contrast, no group differences were observed on the cognitive processing measures used in the current study. A key finding was that musicians outperformed non-musicians in identifying embedded triplets in the aSL but not the vSL task. The musician group showed a triplet onset effect in the ERPs recorded during both online aSL and vSL tasks whereas the non-musician group did not show these effects. A moderate correlation was observed between the music scores and behavioural scores on the aSL task. These findings are further discussed in the context of related studies in the following sections.

4.4.1. Auditory Processing

The results of the present study showed that children who had received musical training for at least 1.5 years or more (mean 3.9 years) outperformed their untrained counterparts in areas closely related to music: melody discrimination, rhythm discrimination, and frequency discrimination. Importantly, these differences were observed even when the variation due to non-verbal IQ was taken into account using a MANCOVA. These findings lend support to previous research in which musical training was seen to enhance melodic and rhythm discrimination performance in children (Flohr, 1981, Morrongiello et al., 1989). Despite controlling for several extraneous variables, some unexplored, pre-existing differences could also account for the pattern of results. For instance, it is possible that children with higher musical abilities would have been more likely to be enrolled for music lessons. However, longitudinal studies in children further support the notion that musical training enhances melody and rhythm discrimination in children (Hyde et al., 2009). Hyde et al. (2009) compared melody and rhythm discrimination skills in 6-year old children, before and after fifteen months of instrumental training, using similar tests as the current study. Children with music training showed greater improvements in these skills as opposed to the control group of untrained children. Although it was not part of our primary aim, our findings also show that the Musical Ear Test (Wallentin et al., 2010) may be used to distinguish between groups of musically trained and untrained children.

Similar to the current study, another cross-sectional study also reported significantly smaller frequency discrimination thresholds for 1 kHz stimuli in children with more than a year of music experience as compared to children with no musical training (Banai et al., 2013). In fact, a longitudinal research has shown that even 6 months of musical training was sufficient to improve pitch processing in 8-year old children (Moreno et al., 2009). Previous studies also report that as the duration of training increases, pitch processing improves. For instance, 8-year old children with 4 years of musical training were significantly better than untrained children in detecting pitch changes as small as one-fifth of a tone (Magne et al., 2006). Taken together, these studies provide evidence that musical training, by increasing sensitivity to specific parameters of music such as pitch and rhythm, does enhance children's ability to detect minute changes in melody and rhythm.

4.4.2. Cognitive processing

In the present study, there were no differences found between the groups on measures of cognitive processing. This is in contrast to a previous study that reported improvements in intelligence after 1 year of musical training (Schellenberg, 2004). However, it should be noted that full scale IQ was measured in the studies that showed an improvement in intelligence following musical training (e.g. Schellenberg, 2004, 2011). In our study, owing to constraints in the length of testing sessions, we administered a screening test only which provided us with an estimate of non-verbal IQ (TONI). Nevertheless, our findings are consistent with several other studies in which associations between intelligence and musical training were not observed (Hyde et al., 2009, Moreno et al., 2009, François et al., 2013).

Unlike reports of enhanced working memory in adult musicians (Clayton et al., 2016), we did not observe differences in performance of the two groups on digit span tasks. Our findings are supported by other research where musically trained and untrained children performed similarly on measures of executive function and attention (Moreno et al., 2009, Schellenberg, 2011, François et al., 2013). There are two possible explanations for these findings. Firstly, the effects related to musical training may be observed only after a longer period of musical training. Further research in which working memory is assessed during various time intervals of musical training may elucidate the relationship between working memory capacity and musical training. Secondly, musical training may be associated with performance on working memory and executive function tasks that are more challenging than the ones used in the current study. The second explanation is supported by research which showed that short-term musical training (1 hour per day, 5 days per week for 20 days) was associated with improvements in behavioural and ERP measures of executive function tasks such as the go and no-go tasks (Moreno et al., 2011). Overall, it appears that the relationship between musical training and general cognitive abilities is complex and warrants further investigation.

4.4.3. Statistical learning

Our findings add to the growing evidence that musical training is associated with enhancements in extraction of statistical cues. To the best of our knowledge, an investigation of both aSL and vSL in children with music training has not been undertaken previously. Our findings show that enhancements in SL may be limited to the auditory domain (at least for the stimuli used in the present study). These findings are in line with our recent study in adult musicians (Mandikal Vasuki et al., 2016).

Of particular interest are the ERP effects observed in the current study. Unlike a previous adult aSL study in which ERPs were recorded as online measures of segmentation (Abla et al., 2008), we did not observe the presence of N400 response in either group. Instead, we observed a significantly larger N250 response for the initial tone (T1) as compared to the final tone (T3) for of an embedded triplet (Figure 3). Previous SL studies report an enlarged N1 response with a fronto-central distribution in adults and children with musical experience (François et al., 2011, François et al., 2013). However, it should be noted that while Francois and colleagues measured ERP responses *during the test phase* (i.e. after familiarization), we recorded ERPs *during familiarization* as an online measure of segmentation. Previous studies in adults have shown similar enlarged N1 response for word onsets during SL tasks with speech stimuli in adults (Sanders et al., 2003).

There are two possible interpretations of the N250 triplet onset effect in musicians for the aSL stimuli reported in the current study. First, the N250 triplet onset effect could be the manifestation of an emerging N1 triplet onset effect that has been previously described in adults (Abla et al., 2008, Sanders et al., 2009). ERPs to speech sounds in typically developing children, comparable in age to the children in the present study, also show a morphology (P1 and N250) similar to that observed in the current study (see Figure 1; Sharma et al., 2014). Thus, the N250 effect in musician children is analogous to the N1 onset effect seen in adults with higher SL scores (Sanders et al., 2002). Second, N250 could be indexing cognitive processes that arise from prediction of the triplet onsets. For example, N250 effect could be observed if additional attentional resources are allocated to the triplet onset (Hansen et al., 1980, Sanders et al., 2002). Indeed, larger blood oxygenation-level dependent (BOLD) activation signal in adult musicians showed that musicians have an enhanced ability to exert sustained cognitive control (Pallesen et al., 2010). Irrespective of whether the N250 triplet onset effect indexes segmentation or additional cognitive processes, it is clear that this effect proceeds differentially in musicians and non-musicians.

An interesting pattern of ERPs was observed for the visual online task. The morphology of ERPs in visual online task was different from that of the auditory online task. It is notable that although the groups had similar performance on the behavioural vSL task, a triplet onset effect in the N200-P300 region was observed in both groups (Figure 4). This effect was significant for the musician group, whilst a trend was observed for the non-musicians (p=0.07) as well. The distribution of behavioural performance scores on the vSL task (Figure 3) shows that median performance is higher in the musician group. It is plausible that these ERP effects reflect successful online segmentation of the visual stream in both groups and are sensitive to the distribution of scores on the behavioural measure. This is supported by studies which reported that neural measures are more sensitive (or show changes) earlier than behavioural measures. For instance, a previous study by Turk-Browne et al. (2009) measured online vSL using functional magnetic resonance imaging in adults. Larger activations in occipital regions were seen even before participants could discriminate between statistically coherent and random sequences on a behaviour task. Alternatively, it is also possible that non-musicians needed longer periods of familiarization (more than 7 minutes) to show similar ERP effects as the musicians. As ERP investigations into online segmentation of visual stimuli have not been previously undertaken in children with and without musical training, we can only speculate that in the visual domain, ERPs may be more sensitive than behavioural measures for online segmentation in children.

4.4.4. Music skills and behavioural SL

A notable finding from the current study was that individual differences in children's music skills are linked with their capacity for SL, especially in the auditory domain. The music skills measured in this study included melody and rhythm discrimination. Rhythm cues are important for segmentation of continuous speech (Cutler et al., 1992, Echols et al., 1997). Additionally, research in adults has shown that melodic processing measured through melodic expectations is associated with conditional probability (and information content) of sequential auditory events (Pearce et al., 2010). Furthermore, McNealy et al. (2006) have shown that SL measured through functional magnetic resonance imaging (fMRI) activity in adults is correlated with temporal processing skills such as temporal order judgement measured using the Tallal repetition test (Tallal et al., 1973). The association between music skills and aSL in the present study supports the notion put forward by many researchers that auditory skills contribute towards detection of statistical regularities (Emberson et al., 2013, François et al., 2013, François et al., 2014). Future studies can illuminate the matter further by investigating the link between SL in conjunction with other aspects of music perception, such as chord perception or contour perception.

The results from the present study confirm and extend previous literature showing that children who undergo musical training outperform their peers on some auditory processing and cognitive abilities. The correlational design of this study does not allow us to determine definitively whether music causally enhanced SL abilities or whether other variables were responsible for the effects found. It is also possible that children with higher SL skills are predisposed to continue practising music over longer periods of time. The results of the present study, along with a previous longitudinal study (François et al., 2013) provide converging evidence for the association between musical training and aSL.

At present, it is not clear what may be the minimum duration of musical training required to see changes in aSL skills of children undergoing musical training. Future longitudinal studies can help to answer questions whether aSL skills continue to improve as length of musical training increases. A cross-sectional study comprising of children of the same age group but at various stages of musical training (e.g. 3 groups of 10- year old children with 2, 3, and 4 years of musical training) can also help us understand the relationship between duration of musical training and aSL. Such studies can also help to tease apart maturational and musical training effects.

Our findings suggest that musical training can play an important role in facilitating SL abilities, which are a building block of language acquisition. Importantly, the ability to extract statistical cues has been linked with other language abilities such as reading (Arciuli et al., 2012b), comprehension of syntax (Kidd et al., 2016), literacy acquisition in a second language (Frost et al., 2013), and perception of speech in adverse listening conditions (Conway et al., 2010). These findings have potential implications for populations with impaired segmentation and/or implicit learning such as individuals with reading and language impairments (Evans et al., 2009, Jiménez-Fernández et al., 2011, Lum et al., 2013). Further research is needed to explore the interaction between musical training and SL in individuals with dyslexia or specific language impairment. Taken together, these results (while preliminary) support the idea that musical training may be explored as an avenue for improving detection of statistical regularities in individuals with poor SL, thereby helping the overall development of language skills.

4.5. Acknowledgements and conflict of interest

This work was funded by the Australia Awards Scholarships and the HEARing CRC, established and supported under the Cooperative Research Centres Program – Business Australia. This work was also supported by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021). Arciuli was supported by a Future Fellowship granted by the Australian Research Council (FT130101570). Conflicts of interest: none.

4.6. References

Abla D, Katahira K, Okanoya K. On-line assessment of statistical learning by event-related potentials. Journal of Cognitive Neuroscience. 2008;20:952-64.

Abla D, Okanoya K. Visual statistical learning of shape sequences: An ERP study. Neuroscience Research. 2009;64:185-90.

Anaya EM, Pisoni DB, Kronenberger WG. Visual-spatial sequence learning and memory in trained musicians. Psychology of Music. 2016:1-17.

Anvari SH, Trainor LJ, Woodside J, Levy BA. Relations among musical skills, phonological processing, and early reading ability in preschool children. Journal of experimental child psychology. 2002;83:111-30.

Arciuli J, Simpson IC. Statistical learning in typically developing children: the role of age and speed of stimulus presentation. Developmental Science. 2011;14:464-73.

Arciuli J, Simpson IC. Statistical learning is lasting and consistent over time. Neuroscience letters. 2012a;517:133-5.

Arciuli J, Simpson IC. Statistical learning is related to reading ability in children and adults. Cognitive science. 2012b;36:286-304.

Banai K, Ahissar M. Musical experience, auditory perception and reading-related skills in children. PLoS One. 2013;8:e75876.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society Series B (Methodological). 1995:289-300.

Brown L, Sherbenou RJ, Johnsen SK. TONI-4, Test of Nonverbal Intelligence: Pro-Ed; 2010.

Clayton KK, Swaminathan J, Yazdanbakhsh A, Zuk J, Patel AD, Kidd G, Jr. Executive Function, Visual Attention and the Cocktail Party Problem in Musicians and Non-Musicians. PloS one. 2016;11:e0157638.

Conway CM, Bauernschmidt A, Huang SS, Pisoni DB. Implicit statistical learning in language processing: word predictability is the key. Cognition. 2010;114:356-71.

Cutler A, Butterfield S. Rhythmic cues to speech segmentation: Evidence from juncture misperception. JMemL. 1992;31:218-36.

Echols CH, Crowhurst MJ, Childers JB. The perception of rhythmic units in speech by infants and adults. JMemL. 1997;36:202-25.

Emberson LL, Liu R, Zevin JD. Is statistical learning constrained by lower level perceptual organization? Cognition. 2013;128:82-102.

Ettlinger M, Margulis EH, Wong PC. Implicit memory in music and language. The relationship between music and language. 2011:159.

Evans JL, Saffran JR, Robe-Torres K. Statistical learning in children with specific language impairment. Journal of Speech, Language, and Hearing Research. 2009;52:321-35.

Flohr JW. Short-term music instruction and young children's developmental music aptitude. Journal of Research in Music Education. 1981;29:219-23.

Forgeard M, Winner E, Norton A, Schlaug G. Practicing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning. PloS one. 2008;3:e3566.

François C, Chobert J, Besson M, Schön D. Music training for the development of speech segmentation. Cerebral cortex. 2013;23:2038-43.

François C, Schön D. Musical expertise boosts implicit learning of both musical and linguistic structures. Cerebral cortex. 2011;21:2357-65.

François C, Schön D. Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: the role of musical practice. Hearing Research. 2014;308:122-8.
Frost R, Siegelman N, Narkiss A, Afek L. What predicts successful literacy acquisition in a second language? Psychological science. 2013;24:1243-52.
Grouiller F, Vercueil L, Krainik A, Segebarth C, Kahane P, David O. A comparative study of different artefact removal algorithms for EEG signals acquired during functional MRI. Neuroimage. 2007;38:124-37.
Hansen JC, Hillyard SA. Endogeneous brain potentials associated with selective auditory attention. Electroencephalography and clinical neurophysiology. 1980;49:277-90.
Ho Y-C, Cheung M-C, Chan AS. Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. Neuropsychology. 2003;17:439-50.
Hyde KL, Lerch J, Norton A, Forgeard M, Winner E, Evans AC, et al. The effects of musical training on structural brain development. ANYAS. 2009;1169:182-6.
Jasper H. The 10/20 international electrode system. EEG and Clinical Neurophysiology. 1958;10:371-5.
Jiménez-Fernández G, Vaquero JM, Jiménez L, Defior S. Dyslexic children show deficits in implicit sequence learning, but not in explicit sequence learning or contextual cueing. Annals of dyslexia. 2011;61:85-110.
Jung T-P, Makeig S, Humphries C, Lee T-W, Mckeown MJ, Iragui V, et al. Removing electroencephalographic artifacts by blind source separation. Psychophysiology. 2000;37:163-78.
Kidd E, Arciuli J. Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. CD. 2016;87:184-93.
Levitt H. Transformed up-down methods in psychoacoustics. The Journal of the Acoustical Society of America. 1971;49:467-77.
Lum JA, Ullman MT, Conti-Ramsden G. Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies. Research in developmental disabilities. 2013;34:3460-76.
Magne C, Schön D, Besson M. Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. Journal of Cognitive Neuroscience. 2006;18:199-211.
Mahajan Y, McArthur G. Does combing the scalp reduce scalp electrode impedances? Journal of neuroscience methods. 2010;188:287-9.
Mandikal Vasuki PR, Sharma M, Demuth K, Arciuli J. Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning. Hearing Research. 2016;342:112-23.
Maris E. Randomization tests for ERP topographies and whole spatiotemporal data matrices. Psychophysiology. 2004;41:142-51.
Maris E, Oostenveld R. Nonparametric statistical testing of EEG-and MEG-data. Journal of neuroscience methods. 2007;164:177-90.
McNealy K, Mazziotta JC, Dapretto M. Cracking the language code: Neural mechanisms underlying speech parsing. The Journal of neuroscience. 2006;26:7629-39.
Moreno S, Bialystok E, Barac R, Schellenberg EG, Cepeda NJ, Chau T. Short-term music training enhances verbal intelligence and executive function. Psychological Science. 2011;22:1425-33.

Moreno S, Marques C, Santos A, Santos M, Castro SL, Besson M. Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. Cerebral cortex. 2009;19:712-23.
Morrongiello BA, Roes CL, Donnelly F. Children's perception of musical patterns: Effects of music instruction. Music Perception: An Interdisciplinary Journal. 1989;6:447-62.
Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971;9:97-113.
Oostenveld R, Fries P, Maris E, Schoffelen J-M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational intelligence and neuroscience. 2010;2011.
Pallesen KJ, Brattico E, Bailey CJ, Korvenoja A, Koivisto J, Gjedde A, et al. Cognitive control in auditory working memory is enhanced in musicians. PloS one. 2010;5:e11120.
Paraskevopoulos E, Kuchenbuch A, Herholz SC, Pantev C. Statistical learning effects in musicians and non-musicians: An MEG study. Neuropsychologia. 2012;50:341-9.
Patel AD. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Frontiers in Psychology. 2011;2:142.
Pearce MT, Ruiz MH, Kapasi S, Wiggins GA, Bhattacharya J. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. Neuroimage. 2010;50:302-13.
Peter V, Wong K, Narne VK, Sharma M, Purdy SC, McMahon C. Assessing spectral and temporal processing in children and adults using Temporal Modulation Transfer Function (TMTF), Iterated Ripple Noise (IRN) perception, and Spectral Ripple Discrimination (SRD). Journal of American Academy of Audiology. 2014;25:1-9.
Polich J. Attention, probability, and task demands as determinants of P300 latency from auditory stimuli. Electroencephalography and clinical neurophysiology. 1986;63:251-9.
Rohrmeier M, Rebuschat P. Implicit learning and acquisition of music. Topics in Cognitive Science. 2012;4:525-53.
Romano Bergstrom JC, Howard JH, Howard DV. Enhanced implicit sequence learning in college-age video game players and musicians. Applied Cognitive Psychology. 2012;26:91-6.
Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. Cognition. 1999;70:27-52.
Saffran JR, Newport EL, Aslin RN. Word segmentation: The role of distributional cues. JMemL. 1996;35:606-21.
Sanders LD, Ameral V, Sayles K. Event-related potentials index segmentation of nonsense sounds. Neuropsychologia. 2009;47:1183-6.
Sanders LD, Neville HJ. An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. Cognitive Brain Research. 2003;15:228-40.
Sanders LD, Newport EL, Neville HJ. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nature neuroscience. 2002;5:700-3.
Schellenberg EG. Music lessons enhance IQ. Psychological Science. 2004;15:511-4.
Schellenberg EG. Examining the association between music lessons and intelligence. BJP. 2011;102:283-302.
Schlaug G, Forgeard M, Zhu L, Norton A, Norton A, Winner E. Training-induced neuroplasticity in young children. ANYAS. 2009;1169:205-8.
Semel EM, Wiig EH, Secord W. CELF 4: Clinical Evaluation of Language Fundamentals: Pearson: Psychological Corporation; 2006.

Shahin A, Roberts LE, Trainor LJ. Enhancement of auditory cortical development by musical experience in children. Neuroreport. 2004;15:1917-21.
Sharma M, Purdy SC, Kelly AS. The contribution of speech-evoked cortical auditory evoked potentials to the diagnosis and measurement of intervention outcomes in children with auditory processing disorder. Seminars in Hearing: Thieme Medical Publishers; 2014. p. 051-64.
Shook A, Marian V, Bartolotti J, Schroeder SR. Musical experience influences statistical learning of a novel language. The American Journal of Psychology. 2013;126:95.
Stevens DJ, Arciuli J, Anderson DI. Concurrent movement impairs incidental but not intentional statistical learning. Cognitive science. 2015;39:1081-98.
Strait DL, Kraus N, Parbery-Clark A, Ashley R. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. Hearing Research. 2010;261:22-9.
Tallal P, Piercy M. Defects of non-verbal auditory perception in children with developmental aphasia. 1973.
Tsang CD, Conrad NJ. Music training and reading readiness. Music Perception: An Interdisciplinary Journal. 2011;29:157-63.
Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. Journal of Cognitive Neuroscience. 2009;21:1934-45.
Turner A, Sandford J. A normative study of IVA: Integrated visual and auditory continuous performance test. Annual Convention of the American Psychological Association1995.
Wallentin M, Nielsen AH, Friis-Olivarius M, Vuust C, Vuust P. The Musical Ear Test, a new reliable test for measuring musical competence. Learning and Individual Differences. 2010;20:188-96.
Wallin NL, Merker B. The origins of music: MIT press; 2001.

This page is intentionally left blank

Chapter 5

Individual differences in the ability to perceive speech in noise are related to the capacity for statistical learning

Authors: Pragati Rao **Mandikal Vasuki** [1, 2, 3], Mridula **Sharma** [1, 2], Cailyn **Furze** [1], Joanne **Arciuli**[2, 4]

[1]Department of Linguistics, Australian Hearing Hub, 16 University Avenue, Macquarie University New South Wales 2109, Australia
[2]The HEARing CRC, 550 Swanston Street, Audiology, Hearing and Speech Sciences, The University of Melbourne, Victoria 3010, Australia
[3]ARC Centre of Excellence in Cognition and its Disorders, Level 3, Australian Hearing Hub, 16 University Avenue, Macquarie University NSW 2109
[4]Faculty of Health Sciences, University of Sydney, 75 East St, Lidcombe, 1825, Australia

Corresponding author: pragati.mandikal-vasuki@mq.edu.au

Status of the manuscript: Submitted to Trends in Hearing [Manuscript ID- TIH-16-0059]. This manuscript was re-submitted following blind peer review.

Abstract

Statistical learning (SL) is a form of implicit learning whereby individuals are able to extract statistical regularities from input without conscious awareness. The link between SL and two clinical tasks that assess the perception of speech in noise (Listening in Spatialized Noise-Sentences test: LiSN-S, and Speech Perception in Noise: SPIN) was examined in monolingual, normal hearing adults over two experiments. In Experiment 1, there was no significant correlation between SL and performance on the LiSN-S test after controlling for variance associated with general cognitive abilities. In Experiment 2, we observed a significant positive correlation between SL and speech predictability difference scores obtained using the SPIN test at a low signal-to-noise ratio (-10 dB SNR). These findings, while preliminary, indicate that an individual's ability to perceive speech in challenging listening environments may be related to their capacity for SL, depending on the speech perception task that is used. Future studies require a systematic investigation of other commonly used speech perception tasks to better understand this link.

Keywords

Statistical learning, speech perception in noise, LiSN-S, SPIN

5.1 Introduction

The ability to perceive speech in noise is a necessity for children and adults. Previous studies have supported the role of auditory processing, attention, memory, language, and phonological processes for understanding speech in noise. Another factor, one that has received much less attention, is statistical learning (SL). It is thought that SL is helpful in determining word predictability within language processing (Conway, Karpicke, & Pisoni, 2007; Conway & Pisoni, 2008; Saffran, 2003). Hence, it stands to reason that SL may assist in recognition of speech when listening is challenging. The current study explored whether individual differences in the perception of speech in noise are related to the capacity for SL.

SL is an implicit learning mechanism which identifies statistical regularities in input, often without conscious awareness (see seminal study by Saffran, Aslin, & Newport, 1996). A large body of research has demonstrated that individuals can use this mechanism to learn regularities in auditory (e.g., Saffran, Johnson, Aslin, & Newport, 1999), visual (e.g., Kirkham, Slemmer, & Johnson, 2002), and tactile stimuli (e.g., Conway & Christiansen, 2005). Both adjacent and non-adjacent patterns can be learnt through mere exposure (Gomez, 2002; Newport & Aslin, 2004). These statistical regularities are used by infants to learn word categories and meanings associated with words while acquiring language (Lany, 2014). SL also facilitates the acquisition of phrase-level regularities, which may be then used to identify syntactic regularities (Saffran & Wilson, 2003).

In their seminal study, Saffran, Aslin and Newport (1996) embedded trisyllabic nonwords, also known as triplets, within a continuous speech stream during a familiarization phase. A separate test phase followed familiarization. Consistent with the terminology used in subsequent SL studies that have utilised the same kind of task to assess SL, we refer to this as the embedded triplet paradigm, irrespective of whether auditory, visual, verbal or nonverbal stimuli are used (Arciuli & Simpson, 2012; Kidd & Arciuli, 2016; Siegelman & Frost, 2015;

Stevens, Arciuli, & Anderson, 2015). While the embeeded triplet paradigm is commonly used it is not the only way to assess SL. For example, other studies have used serial reaction time tasks to assess SL (amongst others see Hunt & Aslin, 2001; Kaufman et al., 2010). The serial reaction time tasks assess learning by measuring response speed for grammatical and ungrammatical sequences while the embedded triplet tasks rely on a post-exposure recognition test of familiar vs. novel strings.

It has been demonstrated that individual differences in the capacity for SL are related to performance on language processing tasks (Arciuli & Simpson, 2012; Conway, Bauernschmidt, Huang, & Pisoni, 2010; Kidd, 2012). One study found that individual differences in prediction-based processes tapped by SL mechanisms were associated with the processing of non-adjacent dependencies in natural language in adults (Misyak, Christiansen, & Tomblin, 2010). Misyak and colleagues used a language processing task involving comprehension of subject-relative clauses (e.g., “the reporter that attacked the senator admitted the error”) and complex object-relative clauses (e.g., “the reporter that the senator attacked admitted the error”). The SL task used was a serial reaction time task in which participants had to learn non-adjacent dependencies in auditory stimuli. Interestingly, a significant association was seen only between SL and the language task that was associated with greater processing difficulty; processing of object-relative clauses. That is, individuals with better SL performance tended to read and comprehend complex object-relative clauses more easily.

More recently, Kidd and Arciuli (2016) showed that SL ability, along with working memory and vocabulary, predicted performance on a language comprehension task in children. SL ability was assessed using the embedded triplet paradigm with visual stimuli. The language comprehension task included comprehension of simple actives (e.g., the girl is pushing the boy), passives (e.g., “the boy is being pushed by the girl”), subject relative clauses (e.g., the girl that is pushing the boy), and object-relative clauses (e.g., “the boy that the girl is pushing”). In line with Misyak et al. (2010), SL ability was a significant predictor of performance on language tasks with greater processing difficulties such as comprehension of passives and object relative clauses. This was the case even when other variables such as age and IQ were taken into consideration.

One important language processing task is the ability to perceive speech in challenging listening environments. The role of SL in speech perception was discussed by Harris (1970). Harris noted that in any given language, while it is possible for a number of different combinations of phonemes to be spoken, there is a strong statistical probability that certain sounds will follow one another. For example, in English, phonemic sequences such as “phant” or “vator” are often preceded by “ele” (Saffran et al., 1999). This reflects transitional probability; if X occurs, what are the chances that Y will occur, based on the frequency of those events occurring together. In addition to knowledge about co-occurrence of phonemes, knowledge of probabilities in language can help a listener to predict what word will be spoken next in an utterance (Padó, Crocker, & Keller, 2009). Knowledge of statistical probabilities in sequentially presented input may be especially useful when speech is being perceived in challenging listening conditions, that is, either the speech signal quality is somehow degraded, or speech is accompanied by background noise (e.g., the so-called cocktail party effect where there are multiple speakers that comprise the background noise, and/or other kinds of non-speech background noise).

A study by Conway et al. (2010) explored the relationship between SL and the perception of degraded speech. Normal hearing adults were asked to repeat the final words of high and zero predictability sentences. All sentences were degraded by 6-channel vocoding and presented in quiet. The motivation for choosing 6-channel vocoding was to emulate the listening conditions by cochlear implantees (Conway et al., 2007). It is noteworthy that degraded speech is not the same as speech in the presence of background noise. In degraded (vocoded) speech, envelopes in different frequency channels are extracted and used to modulate a noise carrier. Thus, perception depends on number of channels of frequency information present in the vocoded speech. The speech materials used by Conway et al. (2010) were modified sentences from the speech intelligibility in noise (SPIN) test (Clopper & Pisoni, 2006). This version of the SPIN test has matching keywords for 52 high and low predictability sentences. The sentences were chosen across multiple lists from the original SPIN test and are thus different from the commercially available SPIN test used by audiologists. SL was measured using a modified Simon memory game where learning was measured through improved memory span for statistically consistent, structured visual sequences. The researchers reported a positive correlation between SL and the ability to understand degraded speech. Additional experiments in the same study found that this association remained even after controlling for variance associated with attention, working memory, inhibition, and intelligence. Conway et al. (2010) concluded that SL played an important role in speech perception; assessing and extracting information to improve predictability of words in speech especially during adverse listening conditions.

While speech perception can be assessed using degraded speech (e.g., via vocoding as in Conway et al., 2010), a more naturalistic and commonly used clinical listening task is perception of speech in the presence of background noise. In fact, some previous studies have shown that so called ‘degraded speech’ using 6 spectral channels similar to that used by Conway and colleagues resulted in near perfect sentence recognition in quiet for normal hearing individuals (for a review see Shannon, Fu, & Galvin, 2004). By contrast, perception of speech in the presence of background noise (e.g. multi-talker babble) depends on the ability to use spectro-temporal dips in the masker. In normal hearing individuals, performance decreases as the number of talkers in the babble is increased, with minimum performance observed at N=8 talkers (Simpson & Cooke, 2005). Thus, speech perception in the presence of background noise (such as multi-talker babble which emulates the so-called cocktail party effect) potentially represents a more challenging listening condition than 6-channel vocoding. Moreover, listening to speech in the presence of background noise is a more common experience of listeners in their daily life. Questions regarding (i) whether there is an association between SL and perception of speech in noise and; (ii) whether the association depends on the type of speech perception task remain to be explored. Over the course of two experiments, we examined these associations.

In the current study, we chose tasks that were different from those used by Conway and colleagues. Specifically, we used speech perception tasks that required listeners to perceive speech in the presence of noise (rather than presenting listeners with vocoded speech) and we used the embedded triplet paradigm to assess SL. Demonstrating a link between individual differences in speech perception using speech stimuli in the presence of maskers such as multi-talker babble, and performance on a different SL task, would provide converging evidence for the role of SL in perceiving speech under adverse listening conditions. As everyday listening conditions involve environments where background noise (e.g., competing speech) is present, we chose speech perception tasks that were representative of this condition. Perception of speech in the presence of noise was evaluated using two tests that are readily available and commonly used by audiologists in clinical settings: The Listening in Spatialised Noise test (LiSN-S) and Speech Perception in Noise test (SPIN). These two tests were chosen as they assessed different skills associated with speech perception.

The first test, LiSN-S, taps into the ability to perceive speech as measured via a speech recognition threshold (SRT) where better performance is indicated by lower SRTs. The LiSN-S test provides an estimate of the level of noise required to understand 50% of the speech as determined by adaptively varying signal-to-noise ratios (SNRs). The sentences used in the LiSN-S test were developed following guidelines for construction of speech intelligibility tests for children and had considerable semantic context (e.g., “the boys are watching the game”). Thus, the LiSN-S test provided a measure of speech intelligibility when there was considerable semantic context and the level of noise was adaptively varied.

The second test, SPIN, assesses predictive abilities based on processing of sentences at fixed signal-to-noise ratios (SNRs). Target sentences which varied in the predictability of the last word were mixed with babble noise at different SNRs. The SPIN test contains high predictability (HP) sentences which provide the listener with information from the syntactic, semantic, and prosodic cues in the sentences from two or three ‘pointer words’ that help predict the keyword. The test also contains low-predictability (LP) sentences which do not provide syntactic, semantic, and prosodic cues in the sentences that help predict the keyword. Thus, the SPIN test assessed speech intelligibility when background noise was presented at fixed SNRs and the amount of context in the speech material was varied.

As it is hypothesized that SL is a domain-general learning mechanism, previous studies have investigated the link between SL and speech/language processing tasks using nonverbal SL stimuli (e.g., SL was assessed using nonverbal stimuli by Conway et al., 2010; Kidd & Arciuli, 2016). In line with previous studies, we assessed SL using a nonverbal task.

When speech is presented with noise, listeners rely on probabilistic relations among different units, such as phonemes or words or clauses or other acoustic/linguistic cues, in order to understand target utterances. Since SL is an implicit learning mechanism that identifies regularities in speech/language, we hypothesized that individuals who perform better on the speech in noise tasks would exhibit stronger SL ability.

5.2 Experiment 1

In the first experiment, participants undertook an embedded triplet SL task in the auditory modality which utilised musical tones. Participants completed the LiSN-S test which utilised an SRT that was measured when speech was presented with competing stories spoken by two females that served as multi-talker background noise. The SRT was defined as the SNR that yielded 50% intelligibility. We hypothesized that individuals with better SL abilities would be better able to detect speech at more negative SNRs and therefore have better SRTs on this task. Cognitive abilities relating to inhibition and working memory were also measured so that we could partial out common sources of variance associated with SL and speech perception.

5.2.1 Methods

5.2.1.1 Participants

The study was approved by Macquarie University Human Participants Ethics Committee (Ref: 5201300459) and written consent was received from all participants. Participants were volunteers from the Macquarie University community. Eighteen adults (4 males, mean age 28.9 years, SD 9.3) were recruited. All participants were native English speakers, with English being their only fluent language, had no history of hearing loss, and were not musicians (i.e. had not learnt a musical instrument for more than three years, and had not studied any musical instrument in the last two years). All participants were screened for normal hearing at 20 dB HL. In addition, distortion product otoacoustic emissions (DPOAEs) and contralateral acoustic reflexes were present in all participants indicating normal cochlear and middle-ear functioning.

5.2.1.2 Stimuli

Statistical learning: SL was evaluated using a nonverbal embedded triplet task in the auditory modality. The SL task consisted of two phases: a familiarization phase and a surprise test phase. The SL stimuli were similar to those used in a previous study (Saffran et al., 1999). Eleven pure tones generated through MATLAB (R2013a) were used to create 6 tone triplets (ADB, DFE, GG#A, FCF#, D#ED, CC#D). All tones were from a chromatic scale beginning at middle C (261.6 Hz). The six tone triplets are represented musically in Figure 1. All tones were 550 ms in duration (25 ms rise time and 25 ms fall time).

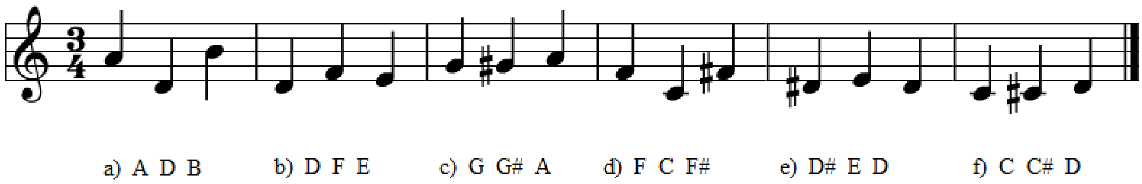

Figure 1. Depiction of six triplets used for familiarization

Familiarization phase: The six tone triplets described above were concatenated randomly to form a continuous tone stream with no gaps. Thus, the triplets were ‘embedded’ within a continuous stream (e.g., ‘CC#DADBGG#AD#ED’). Each triplet was repeated 40 times within the stream. The total length of the stream was 7 minutes and consisted of 240 triplets (40 X 6). There were no acoustic discontinuities between two triplets in a stream or other statistical cues, that is, transitional probabilities were the only indicator of triplet boundaries. The within triplet and across triplet boundary transitional probabilities were calculated using methods published previously (Saffran, Newport, & Aslin, 1996). The across triplet boundary transitional probabilities ranged from 0.04-0.23 (mean 0.1) and the within triplet transitional probabilities ranged from 0.25 to 1 (mean 0.63). As participants listened to the familiarization stream they also performed a cover task to ensure attentiveness. The cover task was an oddball detection task. The oddball stimulus consisted of pure tone of with an F0 of 1319 Hz and there were 40 such oddball stimuli placed randomly at the end of a triplet in the familiarization stream.

Surprise test phase: Participants were unaware of the upcoming SL test phase while completing the familiarization task. The surprise test phase was administered after the completion of the familiarization phase. The test phase consisted of 36 two-alternative forced-choice (2AFC) trials in which an embedded triplet was paired with a novel triplet. Thus, participants were presented with two triplets. One was an embedded triplet that had appeared several times throughout the familiarization task and the other was a novel triplet which was comprised of the same individual tones that had occurred during familiarization but in an order that had not occurred during familiarization. The novel triplets were identical to the triplets described in language 2 by Saffran et al. (1999) (i.e. AC#E, F#G#E, GCD#, C#BA, C#FD, G#BA). The embedded and novel triplets were presented in a counterbalanced order. Participants were asked to indicate which of the two triplets had appeared previously (during familiarization) through a button press response. SL performance was calculated as the percentage of correctly identified embedded triplets.

Speech perception in noise task: We used the Listening in Spatialized Noise Sentences test (LiSN-S; Cameron & Dillon, 2007). Target sentences were presented with two distractor discourses at zero degree azimuth. All speech materials were recorded by female speakers. We administered two conditions of the test: a) target sentence and distractor discourses presented at zero degree azimuth that were spoken by the same individual or same voice-0° condition and b) target and distractors presented at zero degree azimuth which were spoken by different individuals or different voices-0° condition. The target and distractors were presented to both ears simultaneously. The distractors were presented at 55 dB SPL and the level of target sentence was adaptively varied. The participant’s task was to verbally repeat all the words heard in each target sentence. The level of the target sentences was adjusted adaptively by the software to estimate the SRT, which is the level of speech where the participant can repeat at least 50% of the words of a target sentence. Thus, an SRT for each of the two conditions (same voice-0° and different voices-0°) was obtained directly through the LiSN-S software.

Inhibition: Selective attention and inhibition was measured using the Stroop colour word interference task. The test was implemented in Presentation software (www.neurobs.com). Three subtasks were administered – naming the ink color in which the symbol ‘X’ appears, naming of colors printed in black ink, and naming of colors of the printed words. Depending on the subtask, participants were asked to pay attention to ink color or color name. Reaction times were obtained through button press responses for each subtask. Mean reaction times were calculated for all the subtasks. Mean color word interference score was calculated by subtracting the average time needed to complete the first two subtasks from the time needed to complete the third subtask (interference score=Stroop III- [(Stroop I + Stroop II)/2])(Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006).

Digit span tasks: Forwards and backwards digit span subtests from the Clinical Evaluation of Language Fundamentals, 4th edition (CELF-IV; Semel, Wiig, & Secord, 2006) were used. Digits were presented at the rate of one digit per second. Participants were asked to recall the digits in same order (forward span) or reverse order (backward span). Age referenced scaled scores were obtained from raw scores using procedures described in CELF-IV (Australian) norms.

5.2.1.3 Procedure

All participants completed the tasks in a sound treated booth located at the Macquarie University Speech and Hearing Clinic located within the Australian Hearing Hub. The speech perception task was followed by the SL task. The LiSN-S stimuli were administered using a personal computer and Sennheiser HD215 headphones, which were connected directly to the PC. The SL stimuli were through Presentation software (version 16.5, www.neurobs.com) via Etymotic ER 3A insert ear phones at 70 dB SPL.

5.2.2 Results

IBM® SPSS software was used for statistical analyses. During the familiarization phase of the SL task, all participants scored above 90% on the cover task (mean 97.5%). During the surprise test phase of the SL task, the mean percentage of correctly identified embedded triplets was 65.28% (SD 13.75). A one sample t-test revealed that this learning was significantly better than chance (50%) [t (17) = 4.7, $p<0.001$]. On the LiSN-S subtest, one participant scored beyond 3 SD than the group mean on the same voice-0° condition and was excluded. The mean SRT for the same voice-0° condition was -2.6 dB (SD 1.3) while the SRT for different voices-0° was -7.7 dB (SD 3.5). SRT performance for both conditions was within the test norms. The mean Stroop interference score was 154.5 ms (SD 163). Average scores for forward span was 9.5 (SD 2.8) and backward span was 9.5 (SD 2.1).

We performed a Pearson's correlation to examine the relation between SL scores and performance on inhibition tasks as well as between SL scores and performance on memory tasks. The results showed no statistically significant link between the tasks [SL and inhibition: $r = -0.23$, $p=0.35$; SL and forward span: $r=-0.37$, $p=0.13$; SL and backward span: $r=0.09$, $p=0.71$].

Pearson's correlations were used to test our main hypothesis that SL correlates with SRT obtained through the LiSN-S task. Performance on the SL task was not correlated with performance on either of the LiSN-S task conditions (same voice-0°, r=-.34, *p*=0.2, 95% CI= -0.7 to 0.17; different voices-0°, r=-0.36, *p*=0.15, 95% CI= -0.71 to 0.13). Figure 2 shows a scatterplot of SL scores and SRTs obtained for the two LiSN-S tasks. We also computed partial correlations after controlling for variance due to inhibition and memory scores. The results did not achieve conventional levels of significance in either condition (same voice-0°, r=-0.5, *p*=0.07, 95% CI= -0.03 to -0.79; different voices-0°, r=-0.48, *p*=0.07, 95% CI=-0.02 to -0.77). We acknowledge that these results might be described by some as marginally significant.

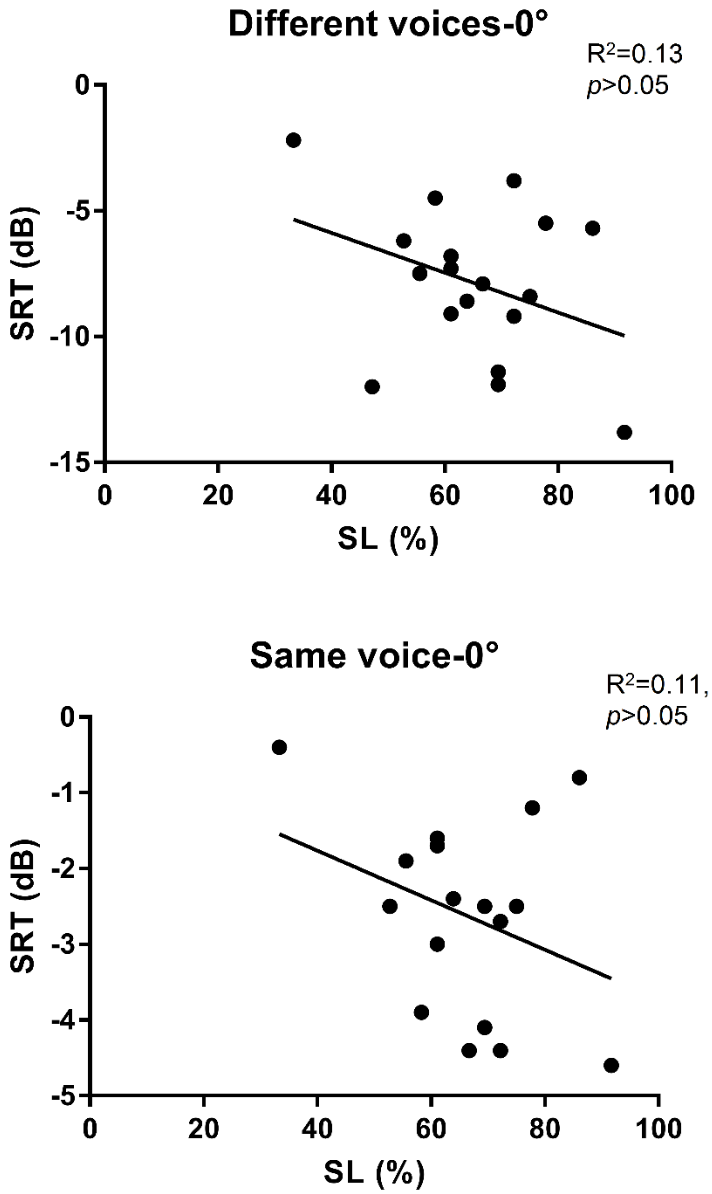

Figure 2. Scatterplot showing percentage of correctly identified triplets for the SL task on the X axis and SRT scores for the LiSN-S test on the Y axis. Top panel =different voices- 0°condition and bottom panel = same voice 0°condition.

5.3 Experiment 2

In Experiment 1, there was no statistically significant relation between SL and speech perception in noise. We did not see strong evidence that the relationship between SL and speech perception is mediated by cognitive abilities in Experiment 1. That is, the association between these variables did not reach conventional levels of significance either before or after partialling out cognitive abilities. As Experiment 1 showed a non-significant result after some measures of cognitive ability were taken into account, we did not pursue assessment of broader cognitive abilities in Experiment 2. Rather, we limited our investigation to assessing the link between performance on an SL task and performance on another clinically available speech perception task.

In Experiment 2, we used a clinical speech perception in noise test that used sentences varying in the predictability of the last word (Speech Perception in Noise; SPIN; Kalikow, Stevens, & Elliott, 1977). Participants' use of context was characterized using a difference score called a predictability difference score (Bilger & Rabinowitz, 1979). Unlike the sentence vocoding used by Conway and colleagues, in the current study, sentence masking was applied using 8-talker babble. Masking using multi-talker speech babble reflects a more realistic listening environment and is more aligned with the kinds of listening tasks used in clinical audiological practice. We hypothesised that there would be a significant correlation between SL ability and predictability difference scores obtained from SPIN test.

5.3.1 Methods

5.3.1.1 Participants

Recruitment, ethics approval and inclusion criteria for participants was identical to that described in Experiment 1. Thirty adults (5 males, mean age 21.3 years, SD 4.6) were recruited. All participants were screened for normal hearing using procedures described in Experiment 1. Written consent was obtained from all participants.

5.3.1.2 Stimuli

Statistical learning: SL was evaluated in the auditory modality using the same task as that used in Experiment 1.

Speech perception in noise task: The speech perception test included in Experiment 1 contained a number of practice trials. In the same vein, we included a practice task in Experiment 2. The practice task was conducted using BKB (Bamford-Kowal-Bench) sentence lists, containing 16 sentences. Speech was presented at 60 dB HL, while multi-talker babble noise (8 talkers) was presented at 50 dB HL, binaurally. Participants were asked to repeat the final word in the sentence. Once participants were familiar with this task, we administered the SPIN test.

The SPIN test consisted of 4 lists -lists A, C, E and G, which consisted of a total of 50 sentences in each list (Kalikow et al., 1977). The speech stimuli were spoken by a female speaker and the noise was 8-talker babble (adult male and female voices). For the SPIN testing, speech and multi-talker babble noise were presented at four SNR levels; 0, -5, -10 and -15 SNR, in order of largest to smallest SNR. This order of presentation is consistent with standard audiological practice where stimuli are presented from high to low SNRs. Noise was presented at 50 dB HL for all sentence lists. Speech intensity was altered to correspond with the chosen SNR level; 50, 45, 40 and 35 dB HL respectively. Speech lists (A, C, E, and G) were allocated randomly to each subject. For the SPIN task, results were scored correct if the participant provided the whole keyword verbally following the presentation of the sentence. See below an example of a highly predictable (HP) and a low predictable (LP) sentence. Participants were asked to repeat the underlined (key) word.

1. HP- The watchdog gave a warning growl.
2. LP- The old man discussed the dive.

At each SNR, a predictability difference score similar to that used by Conway and colleagues was calculated in which performance for highly predictable and low predictable sentences is measured. This was calculated as the difference in the percentage of correctly identified HP sentences minus the percentage of correctly identified LP sentences (see Equation 1). This score was used to determine the participant's use of contextual cues and predictability in different listening conditions. A higher predictability score indicated that a listener was better able to use context.

$$\text{Predictability difference} = HP_{(correct)} - LP_{(correct)}$$

Equation 1: Predictability difference used for SPIN results

5.3.1.3 Procedure

The administration of the SL task was identical to that described under Experiment 1. BKB lists and SPIN lists were presented from the National Acoustic Laboratories (NAL) Compact Disc (CD) (Australia). Both speech tests were presented through headphones (Telephonics TDH-39) via an Interacoustics AC-40 clinical audiometer using a laptop. The speech stimuli were presented binaurally using a 3.5mm splitter 1M to 2F cable.

5.3.2 Results

All statistical analyses were performed using the IBM® SPSS software. During the familiarization phase of the SL task all participants scored above 90% on the cover task (mean 96.7%). The mean percentage of correctly identified embedded triplets during the test phase of the SL task was 59.2% (SD 12.76). A one sample t-test revealed that this learning was significantly better than chance level of 50% [$t(29) = 3.9$, $p<0.001$]. Examination of raw data for the SPIN test showed a floor effect when subjects were tested at -15 SNR with 24 out of 30 participants scoring 0 on both HP and LP conditions. Therefore, these results were not considered for further analysis.

5.3.2.1 Performance across the lists on SPIN

While not the main focus of the paper, closer inspection of the predictability difference scores obtained by administering different SPIN lists revealed a discrepancy in performance by individuals who had been administered List C from the SPIN test. As can be observed from Figure 3, there is a separation in predictability difference results between List C and Lists A, E, and G, which increases with decreasing SNR. Individuals who had been administered List C had the lowest predictability scores. The predictability difference results from other lists are distributed similarly, and do not appear to have any discernible patterns. This was confirmed statistically by testing for a list effect at each SNR using a repeated measures ANOVA. There was a significant effect of list on performance at all SNRs [0 dB SNR: $F(3, 18) = 19.06$, $p < 0.001$; -5 dB SNR: $F(3, 18) = 51.46$, $p < 0.001$; -10 dB SNR: $F(3, 18) = 17.76$, $p < 0.001$]. Pairwise comparisons revealed that performance on list C was significantly different than performance on other lists at the 3 SNRs ($p < 0.05$). However, performance on other lists were similar at the various SNRs (all contrasts $p > 0.05$).

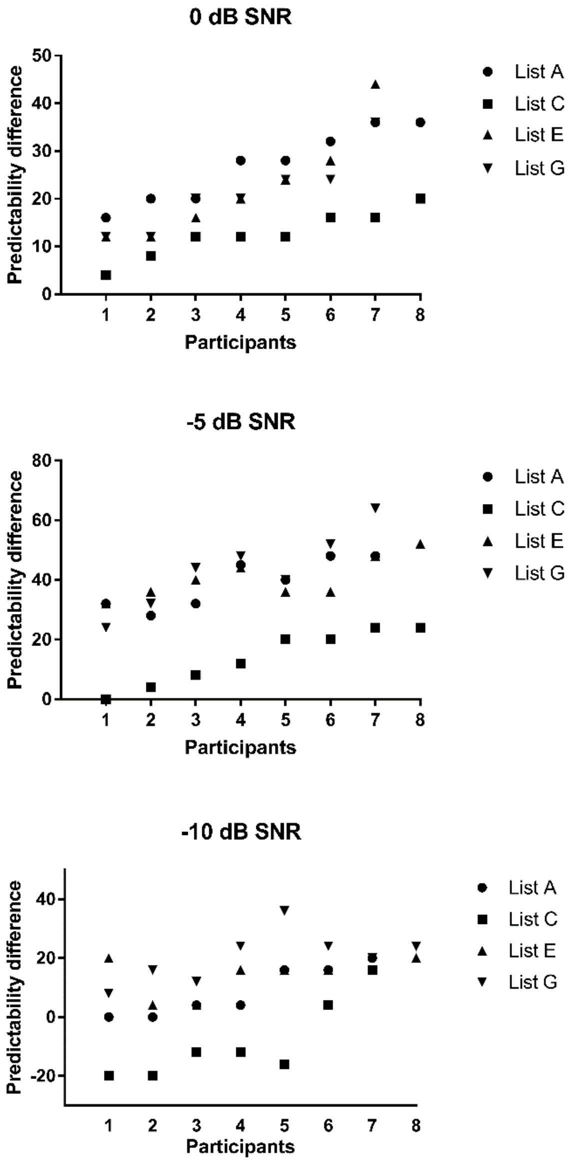

Figure 3. Distribution of predictability difference results of all subjects for SPIN lists presented at different SNRs. The top panel =0 dB SNR, middle panel = -5 dB SNR and bottom panel =-10 dB SNR. Data are represented in ascending order of scores in each SPIN list set (circle = List A; square = List C; upwards triangle = List E; downwards triangle = List G).

Thus, data from all participants who had been administered list C were removed from further analysis. This resulted in removal of data from 8 participants in the 0 dB, -5 dB SNR condition and 7 participants in the -10 dB SNR condition. Table 1 shows the performance on different SNR conditions for the SPIN test for the retained participants.

Table 1. Mean and SDs for performance on SPIN test at all the SNRs tested.

Condition	Measure	Mean	SD
0 dB SNR (N=22)	High predictability (%)	96.5	3.3
	Low predictability (%)	72.9	9.7
	Predictability difference score	23.6	9.1
-5 dB SNR (N=22)	High predictability (%)	83.3	9.2
	Low predictability (%)	42.7	10.4
	Predictability difference score	40.5	9.4
-10 dB SNR (N=23)	High predictability (%)	30.2	10.7
	Low predictability (%)	15.5	6.1
	Predictability difference score	14.7	8.4

Additional analyses were performed where speech recognition threshold (SRT) values were derived by fitting psychometric functions to the speech perception performance in the four levels of SNR (including -15). This process was used to derive SNRs where 50% recognition may be obtained for SPIN sentences (also known as the SRT). The logistic function used for fitting the psychometric function is given in Equation 2. Using this equation, a curve was fit with the percentage correct responses obtained for each participant in the 4 conditions. For an optimal curve fitting, it is preferable to have data points for at least 3 conditions. After fitting of the curve, the SNR on x-axis corresponding to 50% performance on the y-axis, is noted as the SRT value. Due to lack of data points after removal of List C, the SRT values could be derived for only 27 out of 30 participants. The mean SRT for HP sentences was -8.1 dB (SD= 1.2) while the mean SRT for LP sentences was -3.8 dB (SD=1.4).

$$y = \frac{100}{1 + e^{(-4b(SNR-C))}}$$

Equation 2: Logistic function used to derive SRTs for HP and LP sentences in the SPIN test. y represents the percentage correct responses for SPIN sentences, b and C represent the parameters of the logistic function.

5.3.2.2 Link between SL and SPIN

A Pearson's correlation was calculated to assess the relation between SL and the predictability scores of the retained participants. Multiple comparisons were corrected using a false discovery rate technique (Benjamini & Hochberg, 1995). Results showed no correlation between SL and the predictability difference scores of SPIN lists presented at 0 dB SNR (r=-0.03, *p*=0.9, 95% CI= -0.45 to 0.4), and -5 dB SNR (r=0.08, *p*=0.7, 95% CI= -0.35 to 0.49). A significant correlation was found between predictability differences on SPIN lists presented at -10 SNR and SL scores (r=0.48, *p*=0.02, 95% CI= 0.08 to 0.74)[1]. Figure 4 shows a scatterplot of SL scores and predictability differences obtained at different SNRs.

A Pearson's correlation was also calculated to association between the derived SRT values and SL scores. Results showed a significant correlation between SRT values for HP sentences and SL scores (r=-0.41, *p*=0.03, 95% CI= -0.68 to -0.04). A non-significant correlation was found between derived SRT values for LP sentences and SL scores (r=-0.28, *p*=0.15, 95% CI= -0.68 to 0.11).

[1] This association was also confirmed using a non-parametric Spearman's correlation (r_s=0.51, *p*=0.01).

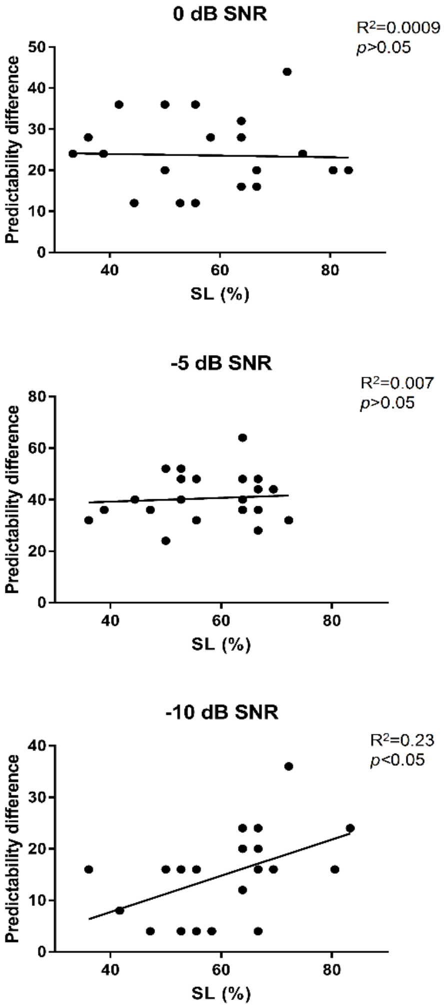

Figure 4. Scatterplot showing percentage of correctly identified triplets for the SL task on the X axis and predictability difference scores for SPIN test on the Y axis. Top panel =0 dB SNR, middle panel = -5 dB SNR and bottom panel =-10 dB SNR.

5.4 General discussion

The aim of the current study was to explore whether individual differences in the ability to perceive speech in noise (using two different speech perception tasks) are related to the capacity for SL. In Experiment 1, our hypothesis was that individuals with better SL would be better at perceiving speech mixed with noise, resulting in a better SRT during the LiSN-S test. In Experiment 2, our hypothesis was that SL would be related to an individual's ability to use context to predict and identify words when speech was presented in noise in the SPIN test.

In Experiment 1 there was no significant correlation between SL and SRTs either before or after variance due to inhibition and memory scores was partialled out. The LiSN-S test measures speech intelligibility at the threshold level, that is, the SNR required to achieve 50% intelligibility. At threshold level, perception of speech is a challenging task and possibly involves the interplay of a variety of cognitive factors including memory. It could be suggested that results from Experiment 1 point to a link between SL and speech perception in noise using the LiSN-S test in that *p* values were closer to conventional levels of significance after partialling out inhibition and memory. This possibility should be investigated in future research with greater statistical power afforded by a larger sample size. Interestingly, our results in Experiment 1 were the same for both subtests of the LiSN-S test even though the two subtests (same voices 0° and different voices 0°) vary in the amount of informational masking produced by the distractors (Glyde et al., 2013). Informational masking is broadly defined as a degradation of the target signal when maskers are highly similar to and/or confusable with the target. In this way masking produces competition at physiological sites beyond the auditory periphery (e.g., “cognitive interference”) (Swaminathan et al., 2015). The same voice distractors condition from the LiSN-S test is considered to be high in informational masking while the different voices distractors condition of the LiSN-S test is lower in informational masking with prominent pitch cues. Future research with a larger sample size, more systematic manipulation of the degree of masking in different conditions, and a larger battery of tests assessing cognitive functions, could help to better understand the relationship between SL and SRT obtained under different masking conditions.

In Experiment 2, our results showed that SL was associated with performance on a speech perception in noise task that drew on predictability. This significant moderate correlation between SL and speech perception was only observed for predictability difference scores of the SPIN test when presented at -10 SNR. It is possible that when listening in challenging acoustic environments, the relative importance of contextual information changes as the difficulty level increases. As such, the relationship between SL and speech perception may be more readily observed in challenging listening conditions. Listeners may benefit from SL mechanisms that assist with prediction because low level auditory processing is compromised by the increasing noise that masks the relevant speech signal. Our findings are in line with previous studies investigating individual differences in SL and language processing where a significant association was seen only for language stimuli that were difficult to process (Kidd & Arciuli, 2016; Misyak et al., 2010).

Additionally, a significant negative correlation was seen between the derived SRT values for HP sentences and the SL scores. The negative r value suggests that individuals who obtained 50% recognition scores at higher SNRs (i.e. more difficult listening situations) also scored higher on the SL task. This result further supports our hypothesis that at challenging listening conditions (such as those SNRs where only 50% recognition is possible), performance on SL task is related with performance on sentence recognition task that relies on heavily on contextual information (i.e. HP sentences). When ambient noise degrades a part of the sentence, a listener may rely on long-term knowledge of sequential probabilities such as word predictability to fill in the missing information. The pattern of results obtained in Experiment 2 suggest that this ability is associated with the more fundamental ability which is used to implicitly learn probabilistic and sequential information that is, SL. It is plausible that superior SL may result in more detailed and robust representation of word order probabilities in spoken language.

In contrast, there were no significant correlations between SL scores and predictability difference scores for SPIN lists presented at 0 dB and -5 dB SNR in Experiment 2. At -5 dB

SNR (mid-SNR) the mean performance for HP sentences was 83.3% while performance for LP sentences was 42.7% (Table 1). Not surprisingly, the increase in level of noise (from 0 to -5 dB SNR) affected performance on LP sentences more than HP sentences. Performance on LP sentences depends on audibility alone as context is not helpful in predicting the keyword. In contrast, recognition of the final word in HP sentences can be achieved through high level cognitive and language processing (contextual cues) and low level auditory processing cues (Lagacé, Jutras, Giguère, & Gagné, 2010). An exploratory Pearson's correlation between SL scores and performance on HP sentences at -5dB SNR showed a significant association ($r=0.46$; $p=0.03$). While speculative, it is possible that at -5dB SNR, the significant correlation between SL and HP indicates that an individual's ability to use context when audibility is compromised may be related to SL. In future research, it would be useful to investigate the relationship between SL and predictability difference scores using finer increments of noise level between 0 and -10 dB SNR.

Previous research examining perception of speech in noise has revealed the importance of a number of different cognitive factors, such as attention, working memory, context, and language skills (Bradlow & Alexander, 2007; Conway et al., 2010; Pichora-Fuller, Schneider, & Daneman, 1995). In Experiment 1, we measured memory (via digit span tasks) and inhibition (via the Stroop task) while Conway et al. (2010) measured other cognitive abilities such as working memory (via forward and backwards digit span), inhibition (via the Stroop Color and Word test) and nonverbal intelligence (via Raven's Standard Progressive Matrices). Although we did not include assessment of cognitive abilities in Experiment 2, it is possible that the relation between SL and sentence recognition in the various SNR conditions of the SPIN test might have been mediated by cognitive processes. Investigation of a broader range of cognitive factors in future research will achieve a better understanding of the relative contributions of these factors when it comes to the relation between SL and perception of speech in noise.

In line with previous studies (Conway et al., 2010; Kidd & Arciuli, 2016), the current study explored SL using a nonverbal SL task. If SL is a domain-general learning mechanism, one would expect a relation between speech perception and SL regardless of whether the learning task utilised verbal or nonverbal stimuli. Our results from Experiment 2 suggest that this is the case and provide converging evidence for the role of SL in speech perception under adverse listening conditions.

The current results have wider implications for current clinical practice in audiology. One of the chief concerns of individuals with hearing impairment is perceiving speech in the presence of noise. There is a high degree of inter subject variability in speech in noise performance among individuals with hearing impairment (Crandell, 1991). Along with investigation of cognitive factors, investigation of individual differences in SL might be helpful in explaining some of the variability seen in the ability to perceive speech in noise in clinical populations.

5.5 Acknowledgements and declaration of conflicting interests

This research was conducted at the Australian Hearing Hub, an initiative of Macquarie University, which brings together Australia’s leading hearing and healthcare organisations to collaborate on research projects. This work was funded by the Australia Awards Scholarships and the HEARing CRC, established and supported under the Cooperative Research Centres Program – an Australian Government Initiative. This work was also supported by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021). Arciuli was supported by a mid-career Future Fellowship granted by the Australian Research Council (FT130101570). The Author(s) declare(s) that there is no conflict of interest.

5.6 References

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*(2), 286-304. doi: 10.1111/j.1551-6709.2011.01200.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.
Bilger, R., & Rabinowitz, W. (1979). Relationships between high-and low-probability SPIN scores. *The Journal of the Acoustical Society of America, 65*(S1), S99-S99.
Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339-2349.
Cameron, S., & Dillon, H. (2007). Development of the listening in spatialized noise-sentences test (LISN-S). *Ear and Hearing, 28*(2), 196-211.
Clopper, C. G., & Pisoni, D. B. (2006). The nationwide speech project: A new corpus of American English dialects. *Speech Communication, 48*(6), 633-644.
Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition: International Journal of Cognitive Science, 114*(3), 356-371. doi: 10.1016/j.cognition.2009.10.009
Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24.
Conway, C. M., Karpicke, J., & Pisoni, D. B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. *Journal of Deaf Studies and Deaf Education, 12*(3), 317-334.
Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Ann N Y Acad Sci, 1145*, 113-131. doi: 10.1196/annals.1416.009
Crandell, C. C. (1991). Individual Differences in Speech Recognition Ability: Implications for Hearing Aid Selection. *Ear and Hearing, 12*(6), 100S-108S.
Glyde, H., Buchholz, J., Dillon, H., Best, V., Hickson, L., & Cameron, S. (2013). The effect of better-ear glimpsing on spatial release from masking. *The Journal of the Acoustical Society of America, 134*(4), 2937-2945.
Gomez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431-436.
Harris, Z. S. (1970). *From phoneme to morpheme*: Springer.
Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General, 130*(4), 658.
Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61*(5), 1337-1351.
Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition: International Journal of Cognitive Science, 116*(3), 321-340. doi: http://dx.doi.org/10.1016/j.cognition.2010.05.011
Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Dev Psychol, 48*(1), 171-184. doi: 10.1037/a0025405
Kidd, E., & Arciuli, J. (2016). Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. *Child Development, 87*(1), 184-193.
Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition: International Journal of Cognitive Science, 83*(2), B35-B42.

Lagacé, J., Jutras, B., Giguère, C., & Gagné, J.-P. (2010). Development of the Test de Phrases dans le Bruit (TPB) Élaboration du Test de phrases dans le bruit (TPB). *CASLPAction!, 34*(4), 261.

Lany, J. (2014). Judging words by their covers and the company they keep: Probabilistic cues support word learning. *Child Development, 85*(4), 1727-1739.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science, 2*(1), 138-153.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127-162.

Padó, U., Crocker, M. W., & Keller, F. (2009). A probabilistic model of semantic plausibility in sentence processing. *Cognitive Science, 33*(5), 794-838.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America, 97*(1), 593-608.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Current Directions in Psychological Science, 12*(4), 110-114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition: International Journal of Cognitive Science, 70*(1), 27-52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606-621.

Saffran, J. R., & Wilson, D. P. (2003). From Syllables to Syntax: Multilevel Statistical Learning by 12-Month-Old Infants. *Infancy, 4*(2), 273-284.

Semel, E. M., Wiig, E. H., & Secord, W. (2006). *CELF 4: Clinical Evaluation of Language Fundamentals*: Pearson: Psychological Corporation.

Shannon, R. V., Fu, Q. J., & Galvin, J., 3rd. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngol Suppl*(552), 50-54.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105-120.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *Journal of the Acoustical Society of America, 118*(5), 2775-2778.

Stevens, D. J., Arciuli, J., & Anderson, D. I. (2015). Concurrent movement impairs incidental but not intentional statistical learning. *Cognitive Science, 39*(5), 1081-1098.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd Jr, G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Sci Rep, 5*.

Van der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2006). The Stroop color-word test: influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment, 13*(1), 62-79. doi: 10.1177/1073191105283427

This page is intentionally left blank

Chapter 6

Chapter summaries

The primary objectives at the outset of this PhD are listed below:

1. Is music training associated with differences in performance on SL tasks in adults?
2. Is performance on SL tasks linked with performance on auditory or cognitive processing measures?
3. Are there differences in online measures of SL in adult musicians as compared to a control group?
4. How do children with music training perform on behavioural measures of SL as compared to a control group?
5. How do children with music training perform on online measures of SL as compared to a control group?

A secondary objective was to investigate the contribution of SL towards real-world functional outcomes such as perception of speech in the presence of noise. This chapter summarises the salient findings of the four manuscripts included in this PhD thesis in the context of these research questions.

6.1. Chapter 2: Musicians' edge: a comparison of auditory processing, cognitive abilities and statistical learning

In Chapter 2, two research questions were addressed: (i) is music training associated with differences in performance on SL tasks in adults?; (ii) is performance on SL tasks linked with performance on auditory or cognitive processing measures? Auditory processing was evaluated by assessing (a) frequency discrimination at 1 kHz, (b) discrimination of iterated rippled noise, (c) detection of amplitude modulation at 4 and 64 Hz, (d) dichotic digits task and (e) perception of speech in the presence of noise. Cognitive processing was evaluated using measures of (a) forwards and backwards digits span, (b) inhibition and (c) sustained auditory and visual attention. The embedded triplet paradigm was used to evaluate aSL and vSL unimodally using pure tone and cartoon-like figures respectively.

Both groups performed above chance level on the two SL tasks. Our findings indicated that musicians outperformed non-musicians on some of the tasks used: frequency discrimination, backwards digit span and aSL. A key finding was that performance on SL tasks was not associated with performance on any of the auditory or cognitive processing tasks used in the current study.

6.2. Chapter 3: Musicians' online performance during auditory and visual statistical learning tasks

In Chapter 3, the third research question was addressed: are there differences in online measures of SL in adult musicians as compared to a control group? Chapter 3 contained two experiments that compared musicians' and non-musicians' online performance on unimodal aSL and vSL tasks described in Chapter 2. Event-related potentials (ERPs) were recorded for tone stimuli (aSL, experiment 1) and cartoon-like figures (vSL, experiment 2). The triplet onset effect was used to measure online segmentation abilities in the two groups. The ERPs to the first stimulus were compared with the ERPs to the third stimulus of an embedded triplet.

In Experiment 1, musicians showed a triplet onset effect in both N1 and N400 regions in the early streams (stream 1). The triplet onset effect was observed at the N400 region for later streams (stream 2 and stream 3) in non-musicians. This effect was observed over the centro-parietal electrodes in both groups. Musicians also outperformed non-musicians on the behavioural aSL task. In experiment 2, the triplet onset effect was observed in all three streams in the two groups in the N1-P2 region over the parieto-occipital electrodes indicating no group differences on the online vSL task. Similarly, no group differences were observed on the behavioural vSL task. In addition, a significant correlation was observed between the ERP waveforms and the behavioural scores in both aSL (experiment 1) and vSL (experiment 2) tasks.

6.3. Chapter 4: Investigation of statistical learning and auditory processing in children with music training

Chapter 4 addressed the fourth and fifth research questions: (iv) how do children with music training perform on behavioural measures of SL as compared to a control group? and (v) how do children with music training perform on online measures of SL as compared to a control group? This chapter contained a cross-sectional study of 50 children who were divided into two groups matched on age, socio-economic status and parents' education level. One group of children (N = 25) were receiving private music instruction and the other group (N = 25) were not receiving any music training. The two groups were measured on SL (two modalities, behavioural and online measures) using the same tasks as in Chapters 2 and 3. Importantly, we investigated individual differences amongst children to assess if higher SL was associated with higher music skills. Due to constraints in length of testing, a smaller battery of tests was used to assess auditory and cognitive processing tasks as compared with Chapter 2. Auditory processing was evaluated by assessing (a) music skills (melody and rhythm discrimination), (b) frequency discrimination at 1 kHz and (c) dichotic digits task. Cognitive processing was assessed using (a) screening test of non-verbal IQ, (b) forwards and backwards digits span and (c) sustained auditory and visual attention.

Significant group differences were observed in measures of auditory processing closely related to music: melody discrimination, rhythm discrimination and frequency discrimination. Using similar analysis techniques as those in Chapter 3, the triplet onset effect was considered a measure of online segmentation. Both groups performed above chance level on the behavioural aSL and vSL tasks. Children trained in music outperformed the untrained children in both behavioural and online measures of aSL similar to the adult musicians (Chapters 2 and 3). In contrast to the findings in Chapter 3, even though the two groups had similar performance on the behavioural vSL task, a triplet onset effect was observed only for the musically trained group. Another notable finding was that individual differences in music skills were significantly correlated with performance on the SL tasks.

6.4. Chapter 5: Individual differences in perception of speech in noise are related to the capacity for SL

The main aim of Chapter 5 was to address the secondary objective that is, to explore the role of SL in every day listening situations such as perception of speech in the presence of noise. Two experiments were reported where the association between SL and speech perception in the presence of noise was examined using two commonly used speech-in-noise tasks in audiology clinics. Learning of conditional statistics in the auditory modality (aSL) was assessed using the same task as that described in Chapter 2. In Experiment 1, the Listening in Spatialized Noise–Sentences (LiSN-S) test was used to assess speech perception. Results showed a non-significant association between aSL and performance on the two subtests of the LiSN-S test, once variation in memory and inhibition were controlled. In Experiment 2, speech perception was evaluated using the Speech Perception in Noise test (SPIN), a clinical task that tapped into predictive abilities based on context. An association between performance on the SL and SPIN tasks was found at −10 dB SNR. Speech recognition threshold (SRT) values were derived by fitting psychometric functions to the four conditions in the SPIN task. A significant association also was seen between performance on the SL task and derived SRTs for highly predictable (HP) sentences.

Chapter 7

General Discussion

The main goal of this PhD research was to study statistical learning (SL), auditory processing, and cognitive processing in individuals with musical experience. This concluding chapter will discuss the key findings in the context of other studies, implications of these findings, and also propose how they may be used towards the design of new experiments.

7.1. Music training and performance on SL tasks in adults and children

Is music training associated with differences in performance on SL tasks in adults?

How do children with music training perform on behavioural measures of SL as compared to a control group?

The results of this PhD research show that musically trained and untrained participants in our studies (adults and children) learnt the statistics of the auditory and visual streams above chance level. Secondly, individuals with music training (both adults and children) perform better than untrained individuals at tasks involving implicit learning of conditional statistics in the auditory modality but not the visual modality. Importantly, this pattern of results is unlike previous studies comparing aSL in musicians and non-musicians where non-musician groups did not learn the statistics of the aSL streams (François, Chobert, Besson, & Schön, 2013; François & Schön, 2011). Thus, better performance of musically trained individuals (adults and children) on our aSL tasks cannot be solely attributed to the fact that the stimuli were too difficult to be learnt by non-musicians. Thirdly, these differences could be seen even in children who had limited musical experience. For instance, the child musicians who had at least 1.5 years of music training outperformed their untrained peers on the aSL task. These findings may be used as a basis to investigate the length of musical training necessary before aSL abilities differ in musically trained and untrained groups.

It could be argued that higher scores on the aSL tasks in musicians may be because the musicians have greater exposure to musical tones than non-musicians. However, it should be noted that the tone triplets used in the present study were not constructed in accordance with the standard rules of musical composition, and did not resemble any typical melodic fragments (e.g., major and minor triads, or familiar three-tone sequences chimes or jingles). Furthermore, research using musical timbres has shown that listeners were able to overcome the perceptual grouping biases (e.g., acoustical similarities/relationships within the Western musical system) and consistently segment the auditory stream by learning the statistical regularities (Tillmann & McAdams, 2004). While it is possible that musicians may be primed for the aSL stimuli, the non-musicians also performed above chance level (Chapter 2 adult non-musicians: mean score 67.2 %; Chapter 4 child non-musicians: mean score 54.7%). This is in contrast to previous studies which used *musical* stimuli (sung language) and reported performance below chance level on the *musical* dimension for non-musically trained children and adults (François et al., 2013; François & Schön, 2011). As the tones were drawn from the Western musical scale in our study, the musicians may be more engaged while performing the aSL tasks. Additionally, it has also been suggested that musicians performed better on SL tasks with tonal stimuli as they might be paying more attention to auditory stimuli in general (Loui, 2012). Thus, factors such as better motivation and engagement could potentially influence the performance on aSL tasks. Future studies may investigate aSL using speech stimuli or unfamiliar musical scales to tease apart the influence of stimuli on aSL. Nevertheless, there are other potential explanations supported by literature which could also help in understanding the enhanced performance on aSL tasks in musicians.

A second explanation for enhanced aSL in musicians may be enhanced encoding of *auditory information* (François & Schön, 2014). Emberson, Liu, and Zevin (2013) reported that learning of higher order structures on the aSL task was dependent on how stimuli were encoded at lower perceptual levels. In other words, participants gained some knowledge of the structure of the familiarization stream by relying on bottom-up processing and encoding of auditory stimuli. As highlighted in Chapter 1, a large body of literature has shown enhanced auditory encoding, that is enhanced encoding of speech as well as music at subcortical and cortical levels, in individuals with music training (e.g., Musacchia, Sams, Skoe, & Kraus, 2007; Musacchia, Strait, & Kraus, 2008; Pantev, Roberts, Schulz, Engelien, & Ross, 2001; Pantev et al., 2003). Enhanced cortical representation for tones of the musical scale was also reported in musicians (Pantev et al., 1998; Schlaug, Jancke, Huang, & Steinmetz, 1995). This enhancement is reflected by larger amplitude and/or shorter latencies for the evoked responses such as frequency following response (FFR), N1 and P2 responses and larger gray matter volume in auditory areas of musicians. These findings suggest a use-dependent functional reorganization in the brain due to musical practice.

Similar findings have also been reported in children, in whom even short-term music training has been shown to enhance cortical representation of sounds. For example, Fujioka, Ross, Kakigi, Pantev, and Trainor (2006) reported an enhanced negative magnetic evoked response (N250m) to violin tones in 4-6 year old children with one year of music training when compared with untrained children. Another longitudinal study showed enhanced pre-attentive processing of syllable duration and voice onset time in 8-10 year old children following one year of music training (Chobert, François, Velay, & Besson, 2014). Remarkably, changes in the neural representation of speech pitch and musical pitch discrimination were also observed after only six months of music training as opposed to no changes due to painting training in 8-year old children (Moreno et al., 2009). Taken together, these findings suggest that music training facilitates the basic biological processes involved in auditory sequence learning, that is better encoding and enhanced pre-attentive discrimination of sounds. These enhancements in encoding of auditory information possibly facilitate the learning of auditory statistical regularities.

A third possible explanation for enhanced aSL in individuals with music training is that both segmentation (aSL) tasks and music processing invoke the same brain networks which are more developed in musicians. This explanation is supported by neuroimaging studies which have identified the brain regions active during aSL tasks (Chapter 1). These brain networks include the left inferior frontal gyrus (IFG), left middle frontal gyrus (MFG), posterior superior temporal gyrus (pSTG) and superior-ventral premotor cortex (svPMC) (Cunillera et al., 2009; McNealy, Mazziotta, & Dapretto, 2006). Similar brain networks (i.e. activity in the IFG/STG) are also implicated in the processing of music, especially melody and rhythm structure (Koelsch et al., 2002; Levitin & Menon, 2003; Tillmann, Janata, & Bharucha, 2003). Another study has shown the activation of left inferior frontal cortex (IFC) during a pure tone aSL task (Abla & Okanoya, 2008). Research has also shown that musicians have increased grey matter volume in the same brain regions (left IFG and Broca’s area) (Sluming et al., 2002). Plastic changes have also been reported in the brain structure of children learning music (Habib & Besson, 2009; Moreno et al., 2009). Taken together, recruitment of similar brain networks for music processing and segmentation tasks as well as potential experience-related plasticity in the brain may underlie enhanced auditory SL in musicians compared with non-musicians.

In contrast, both groups performed at a similar level on the vSL tasks in Chapter 2 and Chapter 4. Thus, our findings indicate an experience-dependent modality-specific enhancement of SL; that is, enhanced SL in one modality may not be associated with enhanced SL in another modality. To the best of our knowledge, both vSL and aSL have not been not been investigated in the same group of musicians and non-musicians in any previous study. Both groups demonstrated significant learning of the statistical regularities embedded in the visual streams on the behavioural task. Similar performance by the two groups on the vSL task may be because neither group had an encoding advantage for these stimuli as these stimuli are not regularly encountered. Further, despite significant learning it is possible that both groups found learning of temporally presented visual sequences equally difficult. This is supported by studies which have shown that tasks involving temporal presentation of visual stimulus elicit lower level of learning compared to tasks involving temporal presentation of auditory stimulus (Welch & Warren, 1980). Perhaps, research in future could study vSL tasks by spatially separating the stimuli since there is some evidence of enhanced visuospatial perception and imagery in musicians (Brochard, Dufour, & Després, 2004). Future longitudinal studies involving music and art training should measure both aSL and vSL to better understand the interactions between training and modality.

7.2. Online measures of SL in children and adults

Are there differences in online measures of SL in adult musicians as compared to a control group?

How do children with music training perform on online measures of SL as compared to a control group?

Behavioural studies provide information about performance but cannot provide direct evidence of the brain mechanisms involved in online segmentation. We used online measures to study SL in individuals with music training. Importantly, we used a non-parametric randomization process to overcome the problem of multiple comparisons. In both adults and children with music training, triplet onsets elicited larger ERPs than subsequent items in the online aSL task. In adults with music training, both the N1 and N400 triplet onset effects were observed, whereas in children with music training only an N250 triplet onset effect was observed. Different morphologies of ERP waveforms in adults and children, that is presence of the N1 and N400 responses in adults but only the N250 response in children, despite the use of identical stimuli, suggest the role of developmental changes in online SL. Future studies could investigate online mechanisms of aSL in children of different age groups. Such a study would help in teasing apart the association between ERP morphological changes and other developmental factors such as maturation of ERPs.

One possible reason for the differential processing of triplet onsets during the aSL task in individuals with music training is that they direct greater attention to initial than to subsequent sounds in a unit such as a triplet (Sanders, Newport, & Neville, 2002). Occurrence of these effects in later streams (in non-musician adults) or absence of these effects (in non-musician children) suggests that although these individuals are able to segment the stream, the process may be slower or more variable in this population. Due to practical constraints in testing time duration, only one stream of aSL familiarization stream was presented while recording ERPs in children. It is possible that non-musician children might show the triplet onset effect if a longer familiarization stream was used (as in the adult study in Chapter 3). Nevertheless, our findings concur with the study by François, Jaillet, Takerkart, and Schön (2014) that musicians are faster at detecting statistical regularities than non-musicians in the online aSL tasks. The absence of triplet onset effects in non-musicians is possibly because they may not be allocating greater attention to word onsets. This explanation is consistent with research where absent or smaller triplet onset effects were reported for participants with low behavioural test scores (Sanders et al., 2002) or when the segmentation task is found difficult (Sanders & Neville, 2003). It is also possible that for the same reason, non-musicians scored lower than musicians on the subsequent behavioural task.

To the best of our knowledge, this is the first study to obtain online measures of vSL in both adults and children with music training. A previous study in non-musician adults by Abla and Okanoya (2009) assessed online vSL with embedded triplet paradigm using geometric shapes and reported results different than our findings as described below. Abla and Okanoya (2009) reported a small and sharp N400 (−1 μV between 400 and 500 ms) for the initial shapes only in the first learning session of high learners. In a paper describing ERP guidelines, it is noted that typical N400 effects are of long duration and do not necessarily appear as a single, clearly defined peak (Duncan et al., 2009). In contrast, the N400 was not observed in the adults or the children in our studies (Chapter 3 & Chapter 4).

One possible explanation for the different ERP effects observed across the two studies (vSL study by Abla & Okanoya, 2009 and the current PhD research) is related to the overall learning as measured by the behavioural vSL task observed in the two studies. The mean score of high learners who showed the N400 effects in the Abla study was 81.8% (minimum > 72.2%). The mean vSL scores in our studies were lower than those obtained in the Abla study (in our studies, for all adult participants: mean score= 54.9%, range= 41.7- 69.4%; all children participants: mean score = 54%, range=38.9- 66.7%). A comparison of these mean scores suggests that the vSL task used in our studies may be more difficult than that used by Abla and Okanoya (2009). Moreover, Abla and Okanaya (2009) used commonly known geometric shapes (e.g., star, circle) for familiarization. It is possible that participants scored higher on the behavioural task in the Abla study because they used sub-vocal rehearsal strategies to form associations between stimuli (Conway & Christiansen, 2005). This previous explanation (i.e. use of sub-vocal rehearsal strategies) is also likely since Abla and colleagues reported the visual N400 effect at the frontal and medial electrode sites. Previous studies have shown that the visual N400 obtained for pictures that violate expectancy (predictability) was observed in the centro-parietal sites and not the frontal/medial sites (e.g., Nigam, Hoffman, &

Simons, 1992). Thus, it is possible that the N400 effect is observed only when higher scores are obtained on the behavioural task (as in the Abla study).

In the online vSL tasks of the current study, participants watched cartoon-like figures flash on the screen one at a time. In both adults and children, we observed enlarged onset potentials (N1-P2) in the parieto-occipital electrodes. The triplet onset effect (N1-P2 region) was observed in both groups of adults and children (non-musician children: $p = 0.07$). The morphology and distribution of these ERP effects is consistent with ERPs elicited by flashing visual images (field flashes) (Hillyard & Kutas, 1983). These findings suggest that the neurophysiological mechanisms underlying online segmentation of visual stimuli may be constrained in the current study by the type of stimuli used and difficulty of the task.

7.3. SL, auditory processing and cognitive processing

Is performance on SL tasks linked with performance on auditory or cognitive processing measures?

It has been proposed, but never empirically tested, that superior SL in individuals with music training may be mediated by enhancements in auditory processing (François et al., 2013; Shook, Marian, Bartolotti, & Schroeder, 2013). In this doctoral research, we examined this association using different measures in adults and children (in Chapter 2 and Chapter 4). In adults, SL was not associated with performance on any spectral and temporal processing tasks. However, auditory processing skills such as music skills (rhythm and melody discrimination) were associated with performance on SL tasks in children. In other words, children with higher musical skills had higher performance on the SL tasks. These findings are consistent with previous research where the role of rhythmic cues in segmentation of continuous speech has been acknowledged (Cutler & Butterfield, 1992; Loukina, Kochanski, Shih, Keane, & Watson, 2009). A better understanding of the relationship between SL and auditory processing may be obtained by studying temporal processing, since previous research has shown that performance on sequential aSL and vSL tasks is affected by temporal cues such as rate of presentation (Conway & Christiansen, 2009; Emberson, Conway, & Christiansen, 2011).

We did not see an association between SL tasks and working memory tasks used in the current study. However, it should be noted that the relationship between working memory and SL is complex. For example, this relationship may be mediated by the type of working memory task used (Janacsek & Nemeth, 2013). Indeed, recent research by Palmer and Mattys (2016) shows that SL may be disrupted by a concurrent memory task such as a two-back task during familiarization. At this stage, it is not clear how working memory resources play a role in SL. It could be the case that attention is required to maintain relevant syllable combinations in working memory via a process of attentional refreshing, as proposed by several clustering models, including PARSER (Perruchet & Vinter, 1998). Another possibility is that further cognitive resources are required to update the information held in working memory during familiarization. Using working memory tasks that tap into the same cognitive processes as SL tasks could clarify the relationship between memory resources and SL.

7.4. Modality and SL

Better performance in aSL than vSL tasks by all participants in this thesis (with and without musical expertise in Chapters 2 and 4) may indicate a bias towards the auditory modality (in musicians and non-musicians) at least for the types of stimuli and the type of SL paradigm used here. There are two possible explanations for this: (a) constraints within the processing modality and (b) stimulus specificity. The first explanation has been explored in previous research studying modality constraints in SL (Conway & Christiansen, 2005, 2006, 2009; Robinson & Sloutsky, 2007, 2013; Saffran, 2002). When auditory or visual stimuli are presented sequentially, performance may be better in the auditory condition. This is probably because the auditory modality is more sensitive to temporal cues whereas the visual modality is more sensitive to spatial cues (Conway & Christiansen, 2005, 2009). It is likely that scores on vSL tasks may improve if stimuli are presented with spatial-temporal cues. Secondly, within each modality, the encoding of stimuli depends on factors such as stimulus complexity and perceptual saliency (e.g., see Fiser & Aslin, 2001, 2002). As a result, stimulus characteristics may produce different patterns of learning in auditory and visual domains. Taking into account both these explanations, Frost, Armstrong, Siegelman, and Christiansen (2015) proposed a novel theoretical perspective on SL. Their model states that domain-general SL principles ‘are constrained to operate in specific modalities, with potential contributions from partially shared brain regions common to learning in different modalities’ (page 123). Further empirical investigations on modality constraints are needed to help us understand whether the impact of musical training is limited to a specific modality.

7.5. SL and speech perception

What is the contribution of SL towards real-world functional outcomes such as perception of speech in the presence of noise?

The results of Chapter 5 show that better performance on SL task was associated with better perception of speech in a challenging listening condition (-10 dB SNR). These results (while preliminary) show that SL is related to an individual’s ability to use context to predict and identify words when speech was presented in noise. Amongst other factors, both high-level cognitive and language processing cues as well as low-level auditory processing cues may assist the perception of speech in tasks where predictability of keywords is changed (e.g., SPIN test). The relative importance of these factors change as listening conditions become more challenging. It is possible that the relationship between SL and speech-in-noise tasks is more evident when individuals rely more on the high level cues since the negative SNR degraded the low-level cues. Importantly, the results of this study paves the way for future studies to examine the association between SL and other speech perception tasks.

7.6. Contribution to the field

Each chapter in the thesis contributes uniquely to the rapidly expanding field of SL. The innovative aspects of this thesis are as follows:

1. systematic investigation of SL, auditory and cognitive processing in musicians and non-musicians (Chapter 2)
2. assessment of visual along with auditory SL (behaviour & online measures) in adult musicians (Chapters 2 & 3)
3. cross-sectional study of music skills and SL in children with music training (Chapter 4)
4. online measures of auditory and visual SL in children (Chapter 4)
5. assessing the link between SL and speech perception using everyday listening tasks such as speech-in-noise tasks (Chapter 5).

Overall, the results of this PhD research show enhancements in aSL were observed in individuals with long-term music training (adult musicians) as well as individuals learning music (child musicians). These differences were also observed in the ERPs recorded during the online aSL task in both groups. Music training was not associated with benefits in extraction of statistical regularities in the visual domain in either adults or children. All participants performed better on the behavioural aSL task than the behavioural vSL task, highlighting the importance of modality and stimulus constraints while measuring SL. In adults, performance on SL tasks was not associated with performance on any auditory or cognitive processing measure used. However, a measure of musical ability was related to performance on SL tasks in children. Lastly, performance on the aSL task was associated with performance on the SPIN task in a challenging SNR.

7.7. Future work

As music consists of a variety of temporal patterns, future research may investigate the link between aSL and different aspects of auditory processing, such as temporal processing, beat perception, and tasks assessing auditory scene analysis. The question of whether enhancements in aSL are limited to the type of stimuli used in the present study may be further explored by studying SL of unfamiliar music scales and artificial languages. A longitudinal study involving music training and measurement of SL, along with a comprehensive auditory processing test battery, could shed further light on the causality and mechanisms underlying enhanced aSL. Finally, studies comparing aSL and vSL in musicians using modality-specific stimuli such as sequential presentation for auditory stimuli and simultaneous/spatially distributed stimuli for vSL may further our understanding of SL mechanisms in individuals with music training.

7.8. Conclusions and implications

The studies included in this thesis show that music training is associated with enhancements in learning of statistical regularities in the auditory domain in adults and children. In addition, equivalent performance on the online and behavioural measures of SL indicates that ERPs can be used as online measures of SL. This would be particularly useful for evaluating SL in clinical populations (e.g., Jeste et al., 2015). The current PhD cannot address the question of whether musical training causally enhanced aSL abilities in musicians. However, the findings add to the converging body of literature that report that music training is associated with enhanced segmentation skills (François et al., 2013; François et al., 2014; François & Schön, 2014; Shook et al., 2013). The findings have potential clinical implications. A growing body of literature has shown that SL is important for tasks as varied as reading (Arciuli & Simpson, 2012), language comprehension tasks (Kidd & Arciuli, 2016; Misyak & Christiansen, 2007; Misyak, Christiansen, & Tomblin, 2010) and perception of speech in adverse listening conditions (Conway, Bauernschmidt, Huang, & Pisoni, 2010). Moreover, poor performance on segmentation abilities has been found in clinical populations such as individuals with dyslexia (Lum, Ullman, & Conti-Ramsden, 2013), specific language impairment (Evans, Saffran, & Robe-Torres, 2009) and hearing impairment (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011). Music training, by fostering SL, may be used to refine existing therapeutic frameworks in clinical populations. Importantly, future studies confirming a causal link between music training and SL are required before considering music as a therapeutic tool.

7.9. References

Abla, D., & Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: a near-infrared spectroscopy study. *Neuropsychologia, 46*(11), 2787-2795.

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research, 64*(2), 185-190. doi: http://dx.doi.org/10.1016/j.neures.2009.02.013

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*(2), 286-304. doi: 10.1111/j.1551-6709.2011.01200.x

Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition, 54*(2), 103-109. doi: http://dx.doi.org/10.1016/S0278-2626(03)00264-1

Chobert, J., François, C., Velay, J.-L., & Besson, M. (2014). Twelve months of active musical training in 8-to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex, 24*(4), 956-967.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: word predictability is the key. *Cognition: International Journal of Cognitive Science, 114*(3), 356-371. doi: 10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities pitting abstract against stimulus-specific representations. *Psychological Science, 17*(10), 905-912.

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology, 21*(4), 561-580.

Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., & Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science, 14*(1), 69-82.

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., . . . Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage, 48*(3), 541-553.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31*(2), 218-236.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., . . . Van Petten, C. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology, 120*(11), 1883-1908.

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *The Quarterly Journal of Experimental Psychology, 64*(5), 1021-1040.

Emberson, L. L., Liu, R., & Zevin, J. D. (2013). Is statistical learning constrained by lower level perceptual organization? *Cognition: International Journal of Cognitive Science, 128*(1), 82-102.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 52*(2), 321-335. doi: 10.1044/1092-4388(2009/07-0189)

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499-504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 458.

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

François, C., Jaillet, F., Takerkart, S., & Schön, D. (2014). Faster Sound Stream Segmentation in Musicians than in Nonmusicians. *PloS one, 9*(7), e101340.

François, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex, 21*(10), 2357-2365. doi: 10.1093/cercor/bhr022

François, C., & Schön, D. (2014). Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: the role of musical practice. *Hearing Research, 308*, 122-128.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences, 19*(3), 117-125.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain, 129*(10), 2593-2608.

Habib, M., & Besson, M. (2009). What do music training and musical experience teach us about brain plasticity? *Music Perception: An Interdisciplinary Journal, 26*(3), 279-285.

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual review of psychology, 34*(1), 33-61.

Janacsek, K., & Nemeth, D. (2013). Implicit sequence learning and working memory: correlated or complicated? *Cortex, 49*(8), 2001-2006.

Jeste, S. S., Kirkham, N., Senturk, D., Hasenstab, K., Sugar, C., Kupelian, C., . . . Norona, A. (2015). Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Developmental Science, 18*(1), 90-105.

Kidd, E., & Arciuli, J. (2016). Individual Differences in Statistical Learning Predict Children's Comprehension of Syntax. *Child Development, 87*(1), 184-193.

Koelsch, S., Gunter, T. C., Cramon, D. Y. v., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: a cortical “language-network” serves the processing of music. *NeuroImage, 17*(2), 956-966.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage, 20*(4), 2142-2152.

Loui, P. (2012). Statistical learning–what can music tell us. *Rebuschat P, Williams J Statistical Learning and Language Acquisition. Mouton de Gruyter*.

Loukina, A., Kochanski, G., Shih, C., Keane, E., & Watson, I. (2009). *Rhythm measures with language-independent segmentation.* Paper presented at the INTERSPEECH.

Lum, J. A., Ullman, M. T., & Conti-Ramsden, G. (2013). Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies. *Research in developmental disabilities, 34*(10), 3460-3476.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *The Journal of Neuroscience, 26*(29), 7629-7639.

Misyak, J. B., & Christiansen, M. H. (2007). *Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language.* Paper presented at the Proceedings of the 29th Annual Cognitive Science Society.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential Expectations: The Role of Prediction-Based Learning in Language. *Topics in Cognitive Science, 2*(1), 138-153.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex, 19*(3), 712-723.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, 104*(40), 15894-15898.

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research, 241*(1), 34-42.

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4*(1), 15-22.

Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology,* 1-12.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*(6678), 811-814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport, 12*(1), 169-174.

Pantev, C., Ross, B., Fujioka, K., Trainor, L. J., Schulte, M., & Schulz, M. (2003). Music and Learning-Induced Cortical Plasticity. *Annals of the New York Academy of Sciences, 999*(1), 438-450.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39*(2), 246-263.

Robinson, C. W., & Sloutsky, V. M. (2007). *Visual statistical learning: Getting some help from the auditory modality*. Paper presented at the Proceedings of the 29th Annual Cognitive Science Society.

Robinson, C. W., & Sloutsky, V. M. (2013). When audition dominates vision: evidence from cross-modal statistical learning. *Experimental Psychology, 60*(2), 113.

Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language, 47*(1), 172-196. doi: 10.1006/jmla.2001.2839

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research, 15*(3), 214-227. doi: http://dx.doi.org/10.1016/S0926-6410(02)00194-5

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature neuroscience, 5*(7), 700-703.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science, 267*(5198), 699.

Shook, A., Marian, V., Bartolotti, J., & Schroeder, S. R. (2013). Musical experience influences statistical learning of a novel language. *The American journal of Psychology, 126*(1), 95.

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.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research, 16*(2), 145-161.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis) similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1131.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638.

Appendix 1
Additional analyses for performance on the oddball task

A1.1 **Chapters 2 and 3**

Musicians and non-musicians obtained high scores on the oddball detection task during the aSL task [musicians: mean=97.6%, SD=3.3; non-musicians: mean =97.6%, SD=1.3]. Independent t-test showed no significant difference between musicians and non-musicians for the performance on the oddball task during the aSL familiarization [t (36) =0.06, p=0.96]. The results confirmed that both groups were equally adept at detection of the oddball stimulus during the aSL task.

A multiple regression with backwards elimination was conducted to see if musicianship and performance on the oddball task predicted the performance on the aSL task. The analysis shows that performance on the oddball task did not significantly predict the performance on the aSL task (β = -0.04, t (33) = -0.3, $p > 0.05$), however musicianship did significantly predict performance on the aSL task (β =-0.5, t (33) = -3.2, $p < 0.05$). The final model was reached with performance on oddball task removed in step 1 and therefore consisted just of musicianship ($F (2, 32) = 10.7$, $p < 0.05$, $R^2 = 0.25$, $R^2_{Adjusted} = 0.23$).

The results of the statistical analyses confirmed that: a) musicians did not have an improved ability to detect the oddball stimulus and b) musicians' advantage for the aSL task was not due their improved ability to detect the oddball stimulus.

A1.1 **Chapter 4**

Similar statistical analyses as above were conducted to confirm whether there were any differences in performance on the oddball detection task in the two groups of children (musicians and non-musicians).

Musicians and non-musicians obtained high scores on the oddball detection task during the aSL task [musicians: mean=93.6%, SD=4.2; non-musicians: mean =91.7%, SD=8.1]. Independent t-test showed there was no significant difference between musicians and non-musicians for the performance on the oddball task during the aSL familiarization [$t(46)=1.04$, $p=0.3$]. The results confirmed that both groups were equally adept at detection of the oddball stimulus during the aSL task.

As with the adult study above, a multiple regression with backwards elimination was conducted to see if musicianship and performance on the oddball task predicted the performance on the aSL task. The analysis shows that performance on the oddball task did not significantly predict performance on the aSL task ($\beta = 0.12$, $t(47) = 0.9$, $p > 0.05$) however, musicianship did significantly predict performance on the aSL task ($\beta = -0.42$, $t(47) = -3.1$, $p < 0.05$). In step 1, performance on the oddball task was eliminated. Thus, the final model showed that only musicianship and not the performance on oddball task explain a significant amount of the variance in the performance on the aSL task ($F(2, 46) = 10.9$, $p < 0.05$, $R^2 = 0.19$, $R^2_{Adjusted} = 0.18$).

The results of the statistical analyses confirmed that: a) musicians did not have an improved ability to detect the oddball stimulus and b) musicians’ advantage for the aSL task was not due their improved ability to detect the oddball stimulus.

Appendix 2

Ethics Approval

MACQUARIE UNIVERSITY

PRAGATI MANDIKAL VASUKI <pragati.mandikal-vasuki@students.mq.edu.au>

Approved- Ethics application- Sharma (Ref No: 5201300459)

Ethics Secretariat <ethics.secretariat@mq.edu.au> 15 August 2013 09:43

To: Dr Mridula Sharma <mridula.sharma@mq.edu.au>
Cc: Professor Katherine Demuth <katherine.demuth@mq.edu.au>, Miss Pragati Rao Mandikal Vasuki
pragati.mandikal-vasuki@students.mq.edu.au

Dear Dr Sharma

RE: "Role of statistical learning and auditory processing in understanding reading and speech perception in noise" (REF: 5201300459)

Thank you for your email dated 29 July 2013 responding to the issues raised by the Macquarie University Human Research Ethics Committee (HREC (Medical Sciences)).

The HREC (Medical Sciences) is fully constituted and operates in accordance with the National Health and Medical Research Council's National Statement on Ethical Conduct in Human Research (2007) (the National Statement) and the CPMP/ICH Note for Guidance on Good Clinical Practice.

I am pleased to advise that the above project has been granted ethical and scientific approval, effective 15 August 2013.

This research meets the requirements of the National Statement which is available at the following website:
http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72.pdf
This letter constitutes ethical approval only.
The following documentation has been reviewed and approved by the HREC

(Medical Sciences):
1. Macquarie University Ethics Application Form (v. 2.1 - Feb 2013)

2. Email correspondence from Dr Sharma addressing HREC's issues (dated 29/07/2013)
3. Macquarie University Participant Information and Consent Form- "Role of statistical learning and auditory processing in understanding reading and speech perception in noise" (v1, dated 29/07/2013)

4. Recruitment Advertisement- "Are you interested in how your brain perceives patters?" (v1, dated 29/07/2013)

5. Advertisement for Sydney's child- "Do you have a child with listening or reading concerns? (6-8 years of age)?" (v1, dated 29/07/2013)

6. Macquarie University Participant Information and Parent/Child Consent Form- "Role of statistical learning and auditory processing in understanding reading and speech perception in noise" (v2, dated 29/07/2013)

7. Following measure to be used:

7.1. Musical Experience Questionnaire (v1.0, dated 29/07/2013)
7.2. Roles of statistical and auditory processing in understanding reading and speech perception in noise (v1, dated 29/07/2013)- Auditory Processing Questionnaire (Adults)
7.3. Roles of statistical and auditory processing in understanding reading and speech perception in noise (v1, dated 29/07/2013)- Auditory Processing Questionnaire (Children)

8. Recruitment Advertisement- "Are you interested in how your brain perceives patterns?" (v1, dated- 29/07/2013)

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement. It is the responsibility of the Principal Investigator to ensure that the protocol complies with the HREC-approval and that a copy of this letter is forwarded to all project personnel.

2. The National Statement sets out that researchers have a "significant responsibility in monitoring, as they are in the best position to observe any adverse events or unexpected outcomes. They should report such events or outcomes promptly to the relevant institution/s and ethical review body/ies, and take prompt steps to deal with any unexpected risks" (5.5.3). Please notify the Committee within 72 hours of any serious adverse events

or Suspected Unexpected Serious Adverse Reactions or of any unforeseen events that affect the continued ethical acceptability of the project.

3. Approval will be for a period of five (5) years subject to the provision of annual reports.

NB. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.
Progress reports and Final Reports are available at the following website:
http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

4. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

5. All amendments to the project must be reviewed and approved by the Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:
http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by Macquarie University. This information is available at the following websites:
http://www.mq.edu.au/policy/
http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/policy

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have ethics approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of ethics approval to an external organisation as evidence that you have approval please do not hesitate to contact the Ethics Secretariat at the address below.

Please retain a copy of this email as this is your official notification of ethics approval.

Yours sincerely
Dr Karolyn White
Director of Research Ethics
Chair, Human Research Ethics Committee (Medical Sciences)

Ethics Secretariat
Research Office
Level 3, Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia
T: +61 2 9850 6848
F: +61 2 9850 4465
http://www.mq.edu.au/research

Appendix 3

Adult questionnaires

General history questionnaires

Auditory Processing Questionnaire (adults) **Date____________**

Your name: ______________________________

Your DOB: ______________________________

Gender Female ☐ Male ☐

Handedness Right ☐ Left ☐ Ambidextrous ☐

Main language at home?
If other than English
Which language, how much (%) and with whom? ______________________________
____
____
____

Address ______________________________
____

Phone ______________________________

Email ______________________________

Does anyone in the family (parent, siblings, aunts, uncles, etc.) have a listening or reading problem?_____ if yes, please describe____
____
____

Educational Information
Have you ever received any special education services? _________
if yes, which Services?____

Do you have any learning difficulties? _____ If yes, please explain___________________________
____
____

Have you ever had any speech-language problems?_______If yes, please explain______________

Do you have or ever had any reading difficulties? Yes ☐ No ☐

If yes, please explain

Does anyone in your family have reading difficulties? Yes ☐ No ☐
If yes, how are they related to you?

Hearing and listening information
Do you have concerns about your hearing? Yes ☐ No ☐
If yes, please provide more information:

Does anyone in your immediate family have a hearing concern? Yes ☐ No ☐
If yes, how are they related to you?

Has anyone raised any concerns about your hearing? Yes ☐ No ☐
If yes, what are their concerns?

Do you have a history of earache, infections or grommets? Yes ☐ No ☐
If yes, since when and how many episodes - please provide as much information as possible:

Do you have a problem listening or understanding?_____ if yes, please describe the problem:____

When was the problem first noticed?____

What treatment did you receive for this problem?

What questions would you like to have answered about your problem?

____
____
____

Behaviours and Characteristics (please tick and elaborate if you have any of the problems outlined here):

____ **Sensitive to loud sound**
____ **Feel confused in noisy places**
____ **Am easily upset by new situations**
____ **Have difficulty following instructions**
____ **Am restless/problems sitting still**
____ **Short attention span**
____ **Difficulties listening in noisy places**

Please provide any additional information:
____
____
____

Up to this point have you received any assistance or therapy for any of your concerns?
Yes ☐ No ☐
____

Do you have any experience of musical/vocal training
Yes ☐ No ☐
If yes, what kind of training and how many years? (Please complete the musical experience questionnaire) Please mention the age at which you started learning music
____
____
____
____

Have you received any assistance at school or speech therapy?

If yes, what kind of assistance or therapy?
____
____
____

Are there any other medical or health concerns?
____
____
____
____

What is the highest level of education you have received?

0-Pre-primary, kindergarten, pre-school 1-Primary

2A/B-Certificate I and II (general enabling, bridging courses)

2C-Certificate I and II (basic vocational)

3A/B-Higher school certificate, university enabling courses, AQF certificate III

3C-AQF statement of attainment

4A/B-Certificate IV

5A-Bachelor, bachelor with honours, master (research and coursework)

5B-Diploma, advanced diploma, graduate certificate, graduate diploma

6-PhD, professional doctorate

Name: ______________________________ Signature: ____________________

Musical Experience Questionnaire

1. Do you have any musical experience? (do you currently play an instrument/sing, have you ever played an instrument/sang, have you ever undergone any musical training/instruction)

Yes ☐ ☐ No ☐

If no, thank you for completing this questionnaire

2. Do you currently play an instrument?

Yes ☐ ☐ No ☐

If no, thank you for completing this questionnaire

3. Instrument (if more than one please list your primary instrument):

____

4. Other instruments:

____

____

5. Genre/style (if more than one please list your primary genre/style):

____

6. Other genres/styles:

____

____

7. How old were you when you first commenced musical training?

____

8. Since commencing musical training have you ceased musical training or practice for any significant period of one year or more?

   Yes ☐ ☐ No ☐

   If yes, please provide details (e.g., how long etc.):

   ____

   ____

   ____

9. Have you been actively training and playing music continuously for the past 10 years?

Yes ☐ ☐ No ☐

10. Do you earn any income from playing/performing music professionally?

    Yes ☐ ☐ No ☐

If no, please go to question 14

11. How many years have you been playing music professionally?

____

12. Is playing/performing music professionally your primary source of income?

Yes ☐ ☐ No ☐

If no, please go to question 14

13. How many years has playing/performing music professionally been your primary source of income?

____

14. Do you earn any other income from the music industry (eg. teaching, recording etc)?

____

____

15. How many hours per week do you spend playing/performing or practicing music?

____

_____

16. Do you or have you ever spoken any tonal languages?

17. How many years of music experience (in total) would you say you have?

____

Appendix 4

Child questionnaires

Part 1: Questionnaire for parents

Some of these questions are about your child and others might be about you/your family. Please read carefully.

Auditory Processing Questionnaire (children) **Date____________**

Child's name: ______________________________

Child's DOB: ______________________________

Gender Female ☐ Male ☐

Handedness Right ☐ Left ☐ Ambidextrous ☐

Main language at home?
If other than English
Which language, how much (%) and with whom?
____
____
____
____

Address ______________________________
____

Phone ______________________________

Email ______________________________

Which grade/year is your child in? ______________________________

Does anyone in the family (parent, siblings, aunts, uncles, etc.) have a listening or reading problem?_____ if yes, please describe____
____
____

Educational Information
Has the child ever received any special education services? _________
if yes, which services?
____

Does the child have any learning/reading difficulties? _____ If yes, please explain
____
____

Did the child ever have any speech-language problems?_______If yes, please explain______________

____

____

Does anyone in the child's family have reading difficulties? Yes ☐ No ☐

If yes, how are they related to the child?

____

Hearing and listening information

Do you have concerns about your child's hearing? Yes ☐ No ☐

If yes, please provide more information:

Does anyone in your immediate family have a hearing concern? Yes ☐ No ☐

If yes, how are they related to the child?

Has anyone raised any concerns about your child's hearing? Yes ☐ No ☐

If yes, what are their concerns?

Does the child have a history of earache, infections or grommets? Yes ☐ No ☐

If yes, since when and how many episodes - please provide as much information as possible:

Does the child have a problem listening or understanding speech?_____ if yes, please describe the problem:____

____

____

____

When was the problem first noticed?

What treatment was received for this problem?

Behaviours and Characteristics (please tick and elaborate if you or the child's teacher has observed any of the problems outlined here):

____ **Sensitive to loud sound**
____ **Feel confused in noisy places**
____ **easily upset by new situations**
____ **Have difficulty following instructions**
____ **restless/problems sitting still**
____ **Short attention span**
____ **Difficulties listening in noisy places**

Please provide any additional information:

Up to this point has the child received any assistance or therapy for any of the concerns?
Yes ☐ No ☐

Does the child have any musical/vocal training
Yes ☐ No ☐
If yes, what kind of training and how many years? (Please complete the musical experience questionnaire) Please mention the age at which the child started learning music

How many hours/week does your child spend reading (school reading or recreational reading-novels, story books)

What extra-curricular activities is your child involved in? List each activity and number of hours/week.

Information about PARENTS (points A-C) (please note – all information collected is strictly confidential; information from these questions will help us see if our respondents are typical of the overall population)

A. **Parents' education** (please provide the highest level of education; choose from the list below and write the number applicable along with details)

0-Pre-primary, kindergarten, pre-school

1-Primary

2A/B-Certificate I and II (general enabling, bridging courses)

2C-Certificate I and II (basic vocational)

3A/B-Higher school certificate, university enabling courses, AQF certificate III

3C-AQF statement of attainment

4A/B-Certificate IV

5A-Bachelor, bachelor with honours, master (research and coursework)

5B-Diploma, advanced diploma, graduate certificate, graduate diploma

6-PhD, professional doctorate

Mother education Code ____________________
Details_____________________________________

Father education
Code_____________________Details______________________________________

B. **Parents' Occupation**

Mother's occupation____

Father's occupation____

C. **Parents' income**

What is the total of all wages/salaries, government benefits, pensions, allowances and other income that each parent usually receives? Click in the box to select your answer. This question appeared in the 2011 Census.

Do not deduct: tax, superannuation contributions, health insurance, amounts salary sacrificed, or any other automatic deductions.

Include the following:

- Pensions/Allowances: family tax benefit, parenting payment, unemployment benefits, Newstart allowance, rent assistance, pensions, student allowances, maintenance (child support), workers' compensation, any other pensions/allowances.
- Other income: interest, dividends, rents (exclude expenses of operation), business/farm income (exclude expenses of operation), income from superannuation, any other income.
- Wages/salaries: regular overtime, commissions and bonuses.

Mother-

$2,000 or more per week ($104,000 or more per year) ☐

$1,500 - $1,999 per week ($78,000 - $103,999 per year) ☐

$1,250 - $1,499 per week ($65,000 - $77,999 per year) ☐

$1,000 - $1,249 per week ($52,000 - $64,999 per year) ☐

$800 - $999 per week ($41,600 - $51,999 per year) ☐

$600 - $799 per week ($31,200 - $41,599 per year) ☐

$400 - $599 per week ($20,800 - $31,199 per year) ☐

$300 - $399 per week ($15,600 - $20,799 per year) ☐

$200 - $299 per week ($10,400 - $15,599 per year) ☐

$1 - $199 per week ($1 - $10,399 per year) ☐

Negative income ☐

Father-

$2,000 or more per week ($104,000 or more per year) ☐

$1,500 - $1,999 per week ($78,000 - $103,999 per year) ☐

$1,250 - $1,499 per week ($65,000 - $77,999 per year) ☐

$1,000 - $1,249 per week ($52,000 - $64,999 per year) ☐

$800 - $999 per week ($41,600 - $51,999 per year) ☐

$600 - $799 per week ($31,200 - $41,599 per year) ☐

$400 - $599 per week ($20,800 - $31,199 per year) ☐

$300 - $399 per week ($15,600 - $20,799 per year) ☐

$200 - $299 per week ($10,400 - $15,599 per year) ☐

$1 - $199 per week ($1 - $10,399 per year) ☐

Negative income ☐

Name: ______________________________ Signature: __________________________

Part 2: Children Musical Experience Questionnaire

Please answer in consultation with your child

1. Does your child have any musical experience? (he/she currently play an instrument/sing, he/she ever played an instrument/sang, has your child undergone any musical training/instruction)?

Yes ☐ ☐ No ☐

2. Is anyone in your household trained in music?

Yes ☐ ☐ No ☐

If yes, what kind of training and how many years (please give details)?

____

____

____

If no, thank you for completing this questionnaire

3. What kind of music does your child learn? (singing, instruments etc.- please give details)

____

4. Other instruments:

____

____

5. Genre/style (if more than one please list your child's primary genre/style):

____

6. Other genres/styles:

____

____

7. How old was your child (age in years and months) when he/she first commenced musical training?

____

8. Does your child take private music lessons? If yes, how many hours/week?

____

9. Since commencing musical training has your child ceased musical training or practice for any significant period of one year or more?

Yes ☐ ☐ No ☐

If yes, please provide details (eg. how long etc):

____

____

____

10. Has your child actively training and playing music continuously for the past 2 years?

Yes ☐ ☐ No ☐

11. Has your child performed in ensembles/choirs (school band, concert band, community groups)?

____

12. How many pieces can your child play/sing? (i.e. play the song/piece with correct notes; maintain the tempo; or play in front of others)? List any or all the apply

13. Is your child learning to read music/can already read music (please provide details)?

14. How many hours per week does your child spend playing/performing or practicing music?

15. Has your child taken any exams (regarding musical expertise)? For example Australian Musical Examinations Board (AMEB) or any such exam conducted by his tutor? Please provide details about the exams and child's performance?

16. Does your child or have you (the parents) ever spoken any tonal languages?

17. How many years of music experience (in total) would you say your child has?

THE END

Appendix 5

Hearing Research 342 (2016)
112–123

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Musicians' edge: A comparison of auditory processing, cognitive abilities and statistical learning

Pragati Rao Mandikal Vasuki [a, b, c, *], Mridula Sharma [a, b], Katherine Demuth [a, b, c], Joanne Arciuli [b, d]

[a] Department of Linguistics, Australian Hearing Hub, 16 University Avenue, Macquarie University, New South Wales, 2109, Australia
[b] The HEARing CRC, 550 Swanston Street, Audiology, Hearing and Speech Sciences, The University of Melbourne, Victoria, 3010, Australia
[c] ARC Centre of Excellence in Cognition and its Disorders, Level 3, Australian Hearing Hub, 16 University Avenue, Macquarie University, New South Wales, 2109, Australia
[d] Faculty of Health Sciences, University of Sydney, 75 East St, Lidcombe, 1825, Australia

ARTICLE INFO

Article history:
Received 29 October 2015
Received in revised form
11 October 2016
Accepted 15 October 2016
Available online 19 October 2016

Keywords:
Musicians
Statistical learning
Auditory processing
Attention
Digit span

ABSTRACT

It has been hypothesized that musical expertise is associated with enhanced auditory processing and cognitive abilities. Recent research has examined the relationship between musicians' advantage and implicit statistical learning skills. In the present study, we assessed a variety of auditory processing skills, cognitive processing skills, and statistical learning (auditory and visual forms) in age-matched musicians (N = 17) and non-musicians (N = 18). Musicians had significantly better performance than non-musicians on frequency discrimination, and backward digit span. A key finding was that musicians had better auditory, but not visual, statistical learning than non-musicians. Performance on the statistical learning tasks was not correlated with performance on auditory and cognitive measures. Musicians' superior performance on auditory (but not visual) statistical learning suggests that musical expertise is associated with an enhanced ability to detect statistical regularities in auditory stimuli.

1. Introduction

Music is a quintessential multisensory activity and musical training involves engagement of multiple neural and cognitive resources. Musicians not only engage in auditory training due to many hours of listening and practising but also multimodal training involving reading and translation of complex symbolic notation into motor activity (Schlaug et al., 2005). Though it is difficult to differentiate abilities that prompt individuals to pursue music training from abilities that may result from music training, some cross-sectional studies comparing musicians with non-musician peers have shown that musicians perform better on certain auditory processing tasks (Fine and Moore, 1993; Micheyl et al., 2006; Zendel and Alain, 2012) and tasks of executive function (Bialystok and DePape, 2009; Zuk et al., 2014). In addition, neurophysiological and brain imaging studies have shown differences in the brain structure and function of musicians and non-musicians. For example, professional musicians have larger grey matter volume in primary motor and somatosensory areas, premotor areas, anterior superior parietal areas, and in the inferior temporal gyrus bilaterally (Gaser & Schlaug, 2003). The structural and functional changes associated with musical expertise are in line with an experience-dependent model of neuroplasticity (Münte et al., 2002). In summary, there is widespread interest in the musicians' advantage in various abilities. The current research was designed to explore whether auditory processing, cognitive processing, and implicit learning of statistical regularities – statistical learning – differ as a function of musical expertise.

1.1. Auditory processing and musical expertise

Some of the most widely investigated abilities associated with musical expertise pertain to auditory processing. Auditory processing is an umbrella term including spectral and temporal processing. Additionally, measures of temporal processing may include envelope processing and fine structure processing. The tests used to measure auditory processing include tests of frequency discrimination, discrimination of iterated rippled noise, detection of

* Corresponding author. Room 602, Level 1, Australian Hearing Hub, Macquarie University, Sydney, 2109, Australia.
E-mail address: pragati.mandikal-vasuki@mq.edu.au (P.R. Mandikal Vasuki).

http://dx.doi.org/10.1016/j.heares.2016.10.008
0378-5955/

amplitude modulation, detection of gaps in noise, dichotic listening tests, and perception of speech in noise.

When considering the relationship between musical expertise and auditory processing, two issues come into play. First, as far as we are aware, no previous study has compared the performance of musicians and non-musicians on auditory processing tasks that address both spectral and temporal processing abilities. Second, the extent to which musical expertise is associated with superior performance on auditory processing tasks remains a controversial issue. On one hand, research has consistently shown that musicians have better frequency discrimination than non-musicians (Kishon-Rabin et al., 2001; Micheyl et al., 2006). On the other hand, contradictory evidence exists for musicians' superior auditory skills for gap detection (Ishii et al., 2006; Rammsayer and Altenmüller, 2006; Zendel and Alain, 2012), perception of speech in noise (Parbery-Clark et al., 2009; Ruggles et al., 2014), dichotic listening tests (Nelson et al., 2003; Špajdel et al., 2007), and other temporal processing skills (Iliadou et al., 2014; Ishii et al., 2006). Given these inconsistencies in the literature, we incorporated a comprehensive battery of auditory processing tasks in the current study.

1.2. Cognitive abilities and musical expertise

Musical expertise might be associated with some aspects of cognitive processing. For instance, it has been reported that adults and children who have undertaken music training have better working memory as measured through digit span and non-word span (Lee et al., 2007). Better performance on executive function measures such as verbal fluency, design fluency and backwards digit span for musicians has also been reported (Zuk et al., 2014). Still, there have been some ambivalent results as to whether or not musical expertise is associated with enhanced cognitive abilities. Using a large battery of tasks assessing cognitive skills such as verbal comprehension, word fluency, mental rotation, closure, perceptual speed, reasoning, and verbal memory, Brandler and Rammsayer (2003) found significant group differences in only two tasks – verbal memory and reasoning. Similar results were reported in another study where musicians were found to have better performance in only two out of the thirteen primary cognitive abilities tested – flexibility of closure and perceptual speed (Helmbold et al., 2005). Additionally, it is unclear whether enhancements are seen only in the auditory modality, such as in auditory attention tasks (Strait and Kraus, 2011), or also in the visual modality, such as divided visual attention tasks (Rodrigues et al., 2007). Given these gaps in the literature, we incorporated a battery of cognitive processing tasks in the current study including a task that assessed both visual and auditory attention.

1.3. Statistical learning and musical expertise

A growing area of interest is musicians' ability to learn statistical regularities implicitly, known as statistical learning (SL). SL was described in a seminal study by Saffran et al. (1996). They showed that participants are able to extract statistical regularities from a continuous stream of individually presented stimuli using information about transitional probabilities. SL has been shown in auditory (aSL) and visual (vSL) modalities. It is thought that SL ability may contribute to key mental activities including musical appreciation, object recognition, and language acquisition (Arciuli and von Koss Torkildsen, 2012; Rohrmeier and Rebuschat, 2012).

Similar to language, music is highly structured and listeners are able to extract regularities from music (François and Schön, 2010). Whilst being unaware of the complex patterns of music, it is possible to implicitly acquire musical knowledge and use this implicit knowledge to form expectancies, and extract regularities from continuous events. Heightened sensitivity to these statistical regularities in continuous speech or non-speech streams may be partly explained by the shared and overlapping cortical regions for music and language (OPERA hypothesis; Patel, 2010, 2011). It could also be argued that musical competence, which is acquired through repeated practise and exposure, primes and sharpens musicians' intuition for performing implicit learning tasks (Rohrmeier and Rebuschat, 2012). In addition, musicians have may have enhanced processing of auditory stimuli (for a detailed review see François and Schön, 2014). For these reasons, it is interesting to study SL in musicians.

Using neurophysiological measures such as electroencephalography and magnetoencephalography, musicians have been shown to have enhancements in neurophysiological indices (such as N100 or N400) in auditory tasks involving the extraction of distributional cues (François et al., 2014; François and Schön, 2011; Paraskevopoulos et al., 2012; Schön and François, 2011). To date, only two studies have demonstrated an advantage for adult musicians in aSL using behavioural indices (Shook et al., 2013 using morse code; Skoe et al., 2013 using tone doublets). A report of improved SL in a longitudinal study of 8- year old children learning music as opposed to a control painting group suggests that there may be a causal link between musical training and SL (François et al., 2013). Although musical expertise has been associated with improved skills in the visual domain, such as enhanced recognition of visual patterns, also known as design learning (Jakobson et al., 2008), an investigation of musicians' vSL has not been undertaken previously.

Any demonstrable musicians' advantage in SL raises further questions as to whether such an advantage is accompanied by advantages in auditory processing or other cognitive skills. Though not directly investigated, enhanced statistical learning of morse code in musicians was attributed to enhanced temporal encoding and/or cognitive skills in musicians (Shook et al., 2013). However, as far as we are aware, this has not been investigated empirically. We used an array of auditory processing tasks and cognitive processing tasks as well as measures of both auditory and visual SL to explore these questions.

1.4. The current study

The primary aims of this research were to ascertain whether musicians and non-musicians perform differently on: a) tests of auditory processing, b) tests of cognition, and c) tests of SL (aSL and vSL). We hypothesized that musical expertise would be associated with better performance on at least some of the auditory processing and cognition measures. We also hypothesized that musicians might outperform non-musicians with regard to aSL but we were not sure what to expect with regard to vSL. Moreover, we were unsure whether performance on SL tasks would be related to performance on the auditory and cognitive tasks.

2. Methods

2.1. Participants

Musicians were defined as adults who started to learn/practise music before the age of 9 years and had at least 10 years of music playing/singing experience. This criterion is based on previous studies with similar populations (Ruggles et al., 2014; Strait et al., 2010). All musicians reported that they still actively practised music. Non-musicians had less than 3 years of musical experience. Eighteen musicians (5 males) and 22 non-musicians (5 males) participated in the study. There was no significant difference in the ages of the musicians (Mdn = 28.0) and non-musicians

(Mdn = 25.0) as assessed by a Mann-Whitney *U* test [U = 163, p = 0.35]. All participants were right handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) and had normal or corrected to normal vision. At the time of data collection, four of the non-musicians were enrolled in the final year of a postgraduate program. Details about participants' music education, instruments played, and educational background are in Table 1.

The participants were recruited through advertisements distributed via group emails or announcement on Facebook community pages. All participants lived in the greater Sydney metropolitan area and had obtained an undergraduate degree. The study was approved by the Macquarie University Human Participants Ethics Committee and written consent was received from all participants. All participants were paid $40 for their participation.

2.2. Tests

All participants completed behavioural testing in a sound treated booth. Participants' hearing was screened at 15 dB HL (all octave frequencies from 0.5 to 8 kHz). Additionally, distortion product otoacoustic emissions (DPOAEs) and contralateral acoustic reflexes were present at clinically normal levels in all participants.

2.2.1. Auditory processing

The auditory processing tests were the dichotic digits test (3 pairs) (Strouse and Wilson, 1999), gaps in noise test (Baker et al., 2008), frequency discrimination of 1 kHz tones, threshold for discrimination of iterated rippled noise (Peter et al., 2014), threshold for detection of 4 Hz and 64 Hz amplitude modulation (Peter et al., 2014), and the listening in spatialized noise sentence test (LiSN-S, Cameron and Dillon, 2007). As test materials and scoring procedures have been published previously, only brief details of the administration and scoring procedures are mentioned in Table 2.

2.2.2. Cognition (memory, inhibition, and attention)

Working memory capacity was evaluated using forwards and backwards digit span subtests from the clinical evaluation of language fundamentals, 4th edition (CELF-IV; Semel et al., 2006). A Stroop colour word test was used to evaluate inhibition and selective attention. The integrated visual auditory continuous performance test was used to measure sustained attention in auditory and visual modalities (IVA-CPT, Turner and Sandford, 1995). A brief description of procedures for the tests is given in Table 2.

2.2.3. Statistical learning

SL was investigated unimodally in the auditory and visual domains. The separate aSL and vSL tasks were designed to be as similar as possible, using the embedded triplet paradigm with a familiarization phase followed by a separate test phase. The aSL and vSL tasks were adapted from Abla et al. (2008), Abla and Okanoya (2009), and Arciuli and Simpson (2011).

Table 1
Details of all participants' musical and educational background.

No	Age of onset (years)	Years of training	Instrument/s	Highest degree obtained
M1	5	48	Piano, recorder, hackbrett	Post graduate
M2	7	28	Piano, flute	Post graduate
M3	4	18	Piano, guitar, vocals, drums, bass guitar	Graduate
M4	7	13	Piano	Post graduate
M5	9	16	Guitar	Graduate
M6	4	22	Piano, violin	Doctorate
M7	9	20	Vocals, guitar	Post graduate
M8	5	20	Piano, vocals (soprano)	Post graduate
M9	5	16	Piano, alto saxophone, flute, vocals	Graduate
M10	7	17	Piano, vocals, guitar, trombone	Post graduate
M11	8	25	Piano, vocals	Doctorate
M12	5	19	Guitar	Post graduate
M13	6	10	Piano, vocals	Graduate
M14	7	15	Piano	Graduate
M15	9	12	Piano, violin	Graduate
M16	7	16	Violin, bass guitar	Graduate
M17	6	52	Piano, percussion	Doctorate
M18	4	25	Piano, bongo, percussion	Graduate
NM1	–	0	–	Post graduate
NM2	6	2	Piano	Doctorate
NM3	11	2	Trombone	Graduate
NM4	–	0	–	Graduate
NM5	–	0	–	Graduate
NM6	–	0	–	Graduate
NM7	10	3	Piano	Doctorate
NM8	–	0	–	Post graduate
NM9	–	0	–	Post graduate
NM10	–	0	–	Graduate
NM11	–	0	–	Graduate
NM12	–	0	–	Graduate
NM13	–	0	–	Doctorate
NM14	–	0	–	Graduate
NM15	–	0	–	Graduate
NM16	–	0	–	Graduate
NM17	–	0	–	Graduate
NM18	–	0	–	Post graduate
NM19	–	0	–	Post graduate
NM20	–	0	–	Graduate
NM21	–	0	–	Graduate
NM22	–	0	–	Graduate

Table 2
Details of various tests used with a brief description of procedure.

Measures	Tests	Procedure
Auditory processing	Frequency discrimination test (1 kHz) (Peter et al., 2014)	*Stimuli:* 1 kHz pure tone served as standard stimulus. The variable (target) stimuli were generated with frequencies ranging from 1001 Hz to 1050 Hz in steps of 1 Hz. All stimuli were 500 ms duration with a ramp of 20 ms. *Procedure:* Thresholds were estimated based on a 3 AFC procedure with a 2-down 1-up tracking method, estimating the 70.7% correct point on the psychometric function (Levitt, 1971). The target signal frequency was reduced after 2 correct responses and was increased after 1 incorrect response. *Response and scoring:* The participant's task was to identify the interval containing the different signal. The step size was initially 5 Hz and was reduced to 1 Hz after two reversals. The arithmetic mean of the last three reversals in a block of 6 was taken as threshold. Log transformation was applied to the thresholds.
	Threshold for discrimination of iterated rippled noise (Peter et al., 2014)	*Stimuli:* The iterated ripple noise (IRN) stimuli were generated using MATLAB 7 using the add-original configuration (IRNO) method described by Yost (1996). The standard stimulus was white noise with zero iteration. The variable IRNO stimuli were created by adding 10 ms delayed copies of white noise with the original noise. The process was repeated 8 times. IRNO were generated at different gain factors (g) that is, attenuation of the delayed repetition relative to the original noise in order to obtain versions of the stimuli with different pitch strengths. The g ranged from 0 to 0.2 in steps of 0.01. All the stimuli were 500 ms in duration with 30 ms rise and fall times. *Procedure:* Thresholds were estimated based on a 3 AFC procedure with a 2-down 1-up tracking method, estimating the 70.7% correct point on psychometric function (Levitt, 1971). In this procedure, the g of the target signal was reduced after 2 correct responses, and was increased after 1 incorrect response. *Response and scoring:* Participants indicated which of the three stimuli had a pitch percept. The step size for the variable stimuli was initially 0.02 and was reduced to 0.01 after two reversals. The arithmetic mean of the last six reversals in a block of 12 was taken as threshold.
	Gaps in noise test (Baker et al., 2008)	*Stimuli:* White noise with duration of 500 ms with ramp of 20 ms was used as the standard stimulus. White noise of 500 ms duration with varying durations of silence inserted in the centre were used as variable stimuli. *Procedure:* Thresholds were estimated based on a 3 AFC procedure with a 2-down 1-up tracking method, estimating the 70.7% correct point on psychometric function (Levitt, 1971). In this procedure, duration of silence in the target signal was reduced after 2 correct responses, and was increased after 1 incorrect response. *Response and scoring:* Participants indicated which of the three stimuli contained a gap. The step size was initially 3 ms and was reduced to 1 ms after two reversals. The arithmetic mean of the last 3 reversals in a block of 6 was taken as threshold.
	Detection of amplitude modulation (Peter et al., 2014)	*Stimuli:* The standard stimulus was a white noise low pass filtered at 20,000 Hz. The standard stimulus was amplitude modulated with varying modulation depths to create variable stimuli. The modulation frequencies used were 4 and 64 Hz. *Procedure:* 3 AFC with a 2-down 1-up tracking method, estimating the 70.7% correct point on psychometric function (Levitt, 1971) was employed to estimate threshold. In this procedure, the depth of modulation in the target signal was reduced after 2 correct responses, and was increased after 1 incorrect response. *Response and scoring:* Participants were asked to indicate which of the three stimuli was not constant. The amplitude modulation thresholds were based on the modulation depth in decibels ($20 \times \log_{10}(m)$). The step size was initially 4 dB and was reduced to 2 dB after two reversals. The arithmetic mean of the last three reversals in a block of 6 was taken as threshold.
	Dichotic digits test (3 pairs) (Strouse and Wilson, 1999)	*Stimuli:* Over the course of 25 trials, three pairs of digits were presented dichotically (total of 6 digits per trial). *Response and scoring:* After each trial, participants were asked to repeat as many of the 6 digits as they could. They were not asked to report digits from each ear separately. The recalled digits were scored according to which ear they were presented and percentage correct scores were obtained. The right ear advantage was calculated by subtracting left ear scores from right ear scores.
	Listening in Spatialised noise (sentences test) (Cameron and Dillon, 2007)	*Stimuli:* The commercially available Listening in Spatialized Noise Sentence test (LiSN-S) was used. Target sentences were mixed with distracter discourse and presented at a simulated 0° azimuth. Two subtests were administered: a) target and distractor discourse were spoken by the same speaker i.e. same voice-0° and b) target and distractor were spoken by different speakers i.e. different voices-0°. *Response and scoring:* Participants were asked to repeat target sentences and ignore the distractor discourse. The level of the target sentence was adjusted adaptively by the software to estimate the speech reception threshold (SRT) for the two subtests. The levels were adjusted in 4 dB steps until the first reversal in performance was recorded and in 2 dB steps thereafter.
Cognition	Forwards and backwards digit span test (Semel et al., 2006)	*Stimuli:* Digits were presented through headphones at the rate of one digit per second. *Response and scoring:* Participants were asked to recall the digits in same order (forward span) or reverse order (backward span). Age referenced scaled scores were obtained from raw scores using procedures described in CELF-IV (Australian) norms.
	Stroop Colour Word test	*Stimuli:* Computerised test which consisted of three subtasks: naming the ink colour in which the symbol 'X' is printed (Stroop I), reading the colour names printed in black ink (Stroop II), and naming the ink colour of the printed words in which ink color and the word differ (e.g., the word 'red' printed in green ink, Stroop III). The test was implemented in Presentation software (www.neurobs.com). *Response and scoring:* Four response buttons with names of colours printed in black ink were used. Participants pressed a button depending on whether they were asked to identify the ink colour or colour name. Reaction times were obtained for each trial. Mean reaction times were calculated for all the subtasks. Mean colour word interference score was calculated by subtracting the average time needed to complete the first two subtasks from the time needed to complete the third subtask (interference score = Stroop III - [(Stroop I + Stroop II)/2]) (Van der Elst et al., 2006).
	Integrated visual auditory continuous performance test (IVA-CPT) (Turner and Sandford, 1995)	*Stimuli:* Computerised test where 500 trials of visual and auditory '1's and '2's were presented pseudorandomly requiring a shift between the two modalities. *Response and scoring:* Over a span of 13 min, the participants were asked to click the mouse every time they saw or heard '1'. The number '2' was to be ignored. Sustained Auditory and Visual attention quotients are automatically generated by the reporting component of the IVA-CPT software, and represent age- and gender-matched population norms.

2.2.3.1. Stimuli used for familiarization.

The stimuli used for creating embedded triplets for the aSL familiarization phase were same as those described in previous studies (Abla et al., 2008; Saffran et al., 1999). Eleven pure tones within the same octave (starting at middle C or 261.6 Hz within a chromatic set) were generated using MATLAB (R2013 a). Tones are labelled according to their musical notation and are depicted in Fig. 1. The tones were 550 ms in duration (25 ms rise time and 25 ms fall time). Three tones were combined in succession to form a triplet. All participants were exposed to a familiarization stream containing 6 triplets (ADB, DFE, GG#A, FCF#, D#ED, CC#D) similar to those described by Saffran et al. (1999).

For the vSL task, stimuli comprised the 11 cartoons (described as 'aliens'), used by Arciuli and Simpson (2011, 2012a). Stimuli are depicted in Fig. 1. Cartoons were scaled such that the maximum height and width of all cartoons were equal. Each cartoon was presented for 550 ms (similar to tones in the aSL task) against a black background. Three cartoon figures were combined in succession to form a triplet. As with the aSL task, all participants were exposed to 6 visual triplets during familiarization.

2.2.3.2. Creation of familiarization stream.

Irrespective of the modality of the SL task, the familiarization streams were constructed in the following manner. Triplets were concatenated in a pseudorandom order to form a continuous stream of stimuli with no gaps between triplets. Thus, triplets were 'embedded' in a continuous stream. This stream consisted of 40 repetitions of each triplet. The familiarization stream was made up of 240 triplets (40 presentations of each of the 6 triplets) and was about 7 min long. Three familiarization streams were created with different pseudorandom ordering of triplets. The triplets were concatenated with two randomization constraints as described in previous studies (Arciuli and Simpson, 2011, 2012b): a) no repeated triplets were allowed and b) no repeated triplet pairs were allowed. The statistical structure for the aSL and vSL familiarization streams was identical. Statistical cues were the only indicator of triplet boundaries as there were no discernible discontinuities at the triplet boundaries.

Statistical cues to the presence of triplets included transitional probabilities (TP) within triplets and across triplet boundaries (Fig. 2). For example, the within-triplet TP for CC#D in Fig. 2 was calculated as the mean of TPs of the doublets 'CC#' and 'C#D'. For the aSL task and the vSL task, the TPs within triplets for all three streams ranged from 0.25 to 1 (mean 0.625). In contrast, the TPs across triplet boundaries were 0.04–0.3 (mean 0.11).

In the aSL familiarization stream, the mean frequencies for the initial, middle, and final tones across all triplets were 341 Hz (SD = 66), 321 Hz (SD = 57), and 370 Hz (SD = 82), respectively. There was no significant difference between the frequencies of the tones at each position within the tone triplets [$F(2, 10) = 1.48$, $p = 0.28$]. The mean frequencies of the tones within a triplet were as follows- ADB = 409.2 Hz, DFE = 324.2 Hz, GG#A = 415.8 Hz, FCF# = 326.9 Hz, D#ED = 311.5 Hz, and CC#D = 277.5 Hz. There was no significant difference between mean pitch interval within versus across triplets (3.1 vs 4.9 half tones in average).

2.2.3.3. Creation of test phase.

For each of the SL tasks, the test phase included 36 two alternative forced choice trials (2 AFC) (Abla and Okanoya, 2009; Saffran et al., 1999). For each trial during the test phase, participants were presented with an embedded triplet and a novel triplet (counterbalanced order of presentation). The novel triplets were made up from the same individual stimuli but had never occurred together as an embedded triplet in the familiarization stream. The novel triplets were formed according to those reported by Saffran et al. (1999). Each individual tone or cartoon within a triplet was presented using the same presentation time (550 ms) as had been used during familiarization. For each trial, the presentation of the embedded versus the novel triplet was separated by a 1-second gap. After the presentation of both triplets, a new screen appeared which prompted participants to identify which of the two triplets had appeared previously (during familiarization).

Before beginning the test phase, four practice trials were performed to ensure that participants waited for presentation of both triplets before responding. The practice trial triplets did not occur during the familiarization or test phases. This was done to avoid interference between the practice and test trials. Verbal feedback was given during the practice trials. No time constraints were imposed during the practice or test trials.

2.3. Procedure

Data were collected over two sessions. Hearing screening, auditory processing tests, and tests of cognition were administered in session 1. The aSL and vSL tests were administered in session 2. Stimuli for assessments in session 1 were presented using a laptop PC through headphones (Telephonics TDH-39) at 50 dB HL via an Interacoustics AC-40 clinical audiometer. The aSL and vSL task stimuli were delivered through Presentation software (Version 16.5, www.neurobs.com). The aSL stimuli were presented via Etymotic ER 3A insert ear phones (at 70 dB SPL). The visual stimuli were presented on a 17-inch CRT monitor placed at 1 m distance from the participant.

2.3.1. Familiarization phase for SL

Participants were exposed to auditory and visual stimulus streams during which they performed an oddball detection task. This task served as a cover task to ensure attentiveness. During the familiarization phase of the aSL task, participants were asked to press a button whenever they heard a pure tone with a frequency of 1319 Hz (not used in any embedded or novel triplets). During the vSL familiarization phase, they were asked to press a button whenever they saw a particular alien figure (not used in any embedded or novel triplets). To ensure that learning was implicit, participants were not given any instructions about the nature of the embedded triplets within the familiarization stream and were not told to learn or remember anything. There were 40 presentations of the oddball stimulus within each familiarization stream for each modality. The oddball stimulus was randomly presented at the end of a triplet.

Half the participants undertook the familiarization phase of the aSL task before the familiarization phase of the vSL task and the order was reversed for the other participants. The entire familiarization phase for both aSL and vSL lasted about 50 min (2 modalities X 3 streams X 7 min). Breaks of up to 5 min were given between presentations of the streams. Participants were informed about the upcoming test phase only after completion of the familiarization phases for both aSL and vSL tasks.

2.3.2. Test phase

The aSL and vSL test phases were administered in the same order as the familiarization phases. For instance, a participant who was first exposed to the aSL familiarization phase completed the aSL test phase task first. The duration of the test phase for each SL task was 5 min.

2.4. Data analysis

The data from 1 musician and 4 non-musicians were excluded for reasons including active middle ear pathology (1 non-musician)

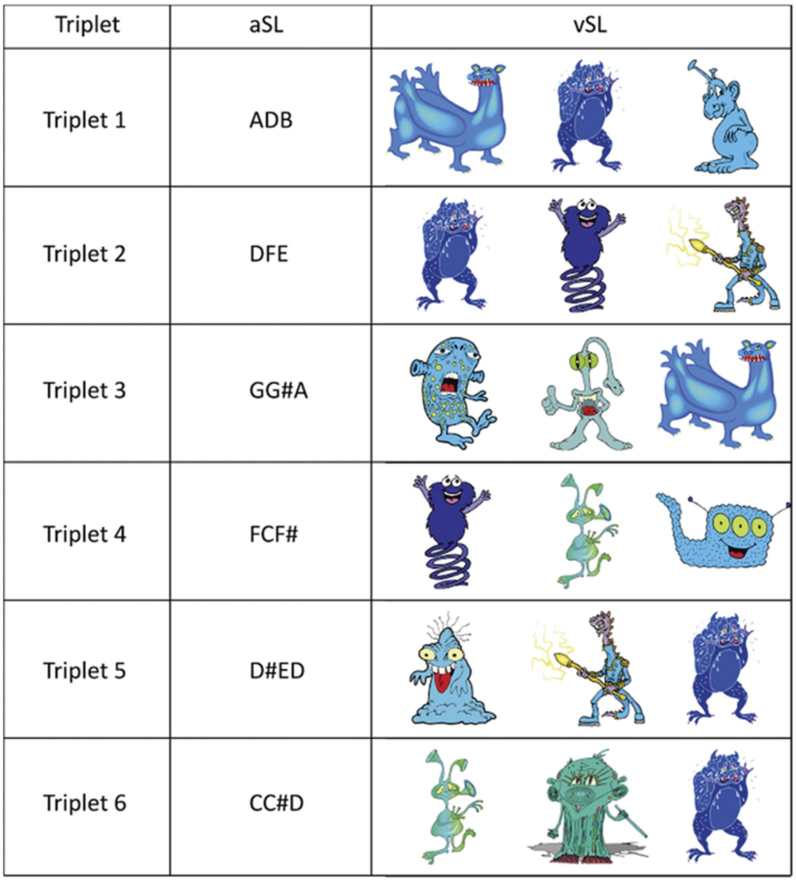

Fig. 1. Triplets used for the aSL and vSL tasks. Stimuli were presented unimodally.

and a score of 0 on the IVA-CPT indicative of attention deficits. Hereafter, any description of the results does not include data for these 5 participants. Thus, data from 17 musicians (4 males, median age 26 years) and 18 non-musicians (4 males, median age 25.5 years) are reported. A Mann-Whitney U test showed no significant difference between the ages of the retained musicians and non-musicians [$U = 143.5$, $p = 0.76$].

Mann-Whitney U tests were used to compare performance on tests of auditory and cognitive processing between the two groups as the data were not normally distributed. Due to multiple comparisons for tests of auditory processing and cognition, we considered p values < 0.01 as significant. A Wilcoxon signed-rank test was used to assess if musical expertise was associated with modality specific enhancements in the attention tasks. Normality for performance on the SL tasks was confirmed using Shapiro-Wilk tests. Pearson's correlations were used to explore associations between SL and tasks of auditory or cognitive processing. All statistical analyses were performed using SPSS version 20 software.

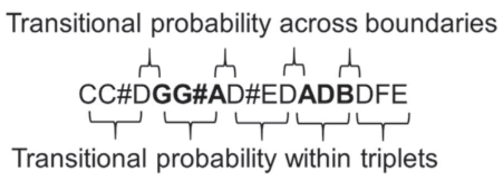

Fig. 2. Calculation of transitional probabilities (TP) for a representational aSL familiarization stream.

3. Results

3.1. Auditory processing

The mean, median, standard deviation, and results from the Mann-Whitney U test for all tests of auditory processing are shown in Table 3. A significant group difference was found for only one auditory processing measure: frequency discrimination. Musicians had better (lower) frequency discrimination thresholds than non-musicians.

3.2. Cognition

Table 4 presents the mean, median, standard deviation and statistics for all the measures of cognition. A significant group difference was found only for backwards digit span, with musicians exhibiting significantly higher scores than non-musicians. There were no group differences on tests of inhibition (Stroop task) and sustained attention. A Wilcoxon signed-rank test showed that the

Table 3
Means, medians, SDs and effect sizes for performance on tests of auditory processing. Significant results are in bold font.

Test	Musicians			Non-musicians			Statistics		
	Mean	Median	SD	Mean	Median	SD	U	p	r
Frequency discrimination test (log transformed)	**0.7**	**0.7**	**0.2**	**1.1**	**1**	**0.8**	**53.5**	**0.001**	**−0.6**
Threshold for discrimination of iterated rippled noise (IRNO gain factor)	0.07	0.05	0.02	0.07	0.07	0.03	123.5	0.3	−0.2
Gaps in noise test (ms)	2.7	2.6	0.4	2.6	2.6	0.4	133.5	0.5	−0.1
Detection of amplitude modulation (4 Hz)- Modulation depth (20 $\log_{10}$ m)	−23.7	−22.7	2.5	−23.3	−24	4.9	149.5	0.9	−0.01
Detection of amplitude modulation (64 Hz)- Modulation depth (20 $\log_{10}$ m)	−14.5	−14.7	3.3	−13.6	−12.7	2.9	121.5	0.3	−0.2
Dichotic digits test − right ear scores (%)	94.8	94.7	4.3	94.2	94.7	4.2	143.5	0.8	−0.4
Dichotic digits test − left ear scores (%)	94.4	94.7	5.3	89.9	89.3	6.5	82	0.02	−0.05
Dichotic digits test (right ear advantage)	0.4	0	5.9	4.4	4.7	5.1	94.5	0.05	−0.3
Listening in Spatialized noise different voices-0° (SRT in dB)	−7.8	−7.4	3.5	−7.9	−7.7	3.1	144.5	0.8	−0.04
Listening in Spatialized noise same voices-0° (SRT in dB)	−2.9	−2.5	2.2	−3.1	−2.5	2.2	139.5	0.7	−0.07

two groups performed at a similar level for auditory and visual attention tasks (musicians: Z = −0.55, p = 0.6; non-musicians: Z = −1.45, p = 0.15). There was no difference between musicians and non-musicians on either auditory or visual attention tasks. Additionally, modality specific enhancements in attention were not observed for musicians.

3.3. Statistical learning

Data from the oddball detection tasks during familiarization were analysed to determine the percentage of successfully identified stimuli. Each participant scored above 80% in the aSL and vSL oddball detection cover tasks. The percentage of correctly identified embedded triplets during the test phase was recorded for musicians and non-musicians. Consistent with previous SL studies (Arciuli and Simpson, 2012a; Conway et al., 2010; Stevens et al., 2015), participants who scored outside the range mean ± 2 SD were excluded from further analyses. This resulted in the exclusion of 4 participants − one non-musician for an excessively low score on the aSL task, one musician for an excessively high score on the vSL, and two non-musicians for excessively low scores on the vSL task. Thus, 34 participants were retained for aSL (17 musicians) and 32 participants for vSL (16 musicians).

Fig. 3 depicts the percentage of correctly identified embedded triplets for retained participants on the aSL and vSL tasks. We performed item analyses to ensure that responding was consistent across the six embedded triplets presented during the test phase. For example, the transitional probabilities were higher within some triplets than others. Results showed that responding was consistent across triplets for both groups. Normal distributions of performance on the aSL and vSL tasks was confirmed for both groups using the Shapiro-Wilk test (all $ps > 0.05$). One sample t-tests revealed that musicians and non-musicians performed significantly above chance on the aSL task [musicians t (16) = 14.9, $p < 0.001$, d = 3.62; non-musicians t (16) = 6.1, $p < 0.01$, d = 1.49]. Likewise, performance on the vSL task was significantly above chance for both groups [musicians t (15) = 2.7, $p < 0.05$, d = 0.67; non-musicians t (15) = 2.9, $p < 0.05$, d = 0.73].

We conducted a 2 × 2 ANOVA to compare performance on the two SL tasks (modality: aSL and vSL) across the two groups (musicians and non-musicians). There was a significant main effect of modality [F (1, 29) = 61.57, $p < 0.005$, partial $\eta^2 = 0.68$], and of group [F (1, 29) = 5.71, $p < 0.05$, partial $\eta^2 = 0.16$], and a significant interaction between group and modality [F (1, 29) = 9.87, $p < 0.01$, partial $\eta^2 = 0.25$]. Pairwise comparisons using t-tests with Bonferroni correction revealed that musicians significantly outperformed non-musicians on the aSL task [t (30) = 3.29, $p < 0.05$, d = 1.13] but not on the vSL task [t (30) = −0.49, $p > 0.05$, d = 0.17].

Scores on the aSL and vSL tasks were not correlated (all participants: r = −0.02, $p > 0.05$; musicians: r = 0.37, $p > 0.05$; non-musicians: r = −0.15. $p > 0.05$). For the musicians, there were no significant associations between SL performance and age of onset of music training (aSL: r = 0.47, $p > 0.05$; vSL: r = 0.34, $p > 0.05$) or between SL and years of music training (aSL: r = −0.23, $p > 0.05$; vSL: r = −0.1, $p > 0.05$).

3.4. Relationship between statistical learning and measures of auditory processing and cognition

Combining the data for both groups, we performed a Pearson's correlational analysis to explore the relationships between SL and measures of auditory processing and cognition. Scores for aSL were not correlated with any auditory processing or cognitive measures except frequency discrimination. There was a moderate negative correlation between performance on aSL and frequency discrimination thresholds (r = −0.43, p = 0.01). To tease apart the influence of musical expertise in this association, Pearson's r was calculated separately for each group. The results revealed no significant relationships (musicians: r = 0.35, p = 0.17; non-musicians: r = −0.45, p = 0.07).[1] vSL was not correlated with any of the measures of auditory processing or cognition. Based on the results of these correlational analyses and to avoid the risk of over-fitting, a follow up multiple-regression analysis was not conducted.

4. Discussion

Our findings indicate that musicians performed better than non-musicians in one auditory and one cognitive task: frequency discrimination and backward digit span. A key finding was that musicians outperformed non-musicians in identifying embedded triplets in the aSL task but not the vSL task. Performance on the SL tasks was not correlated with performance on the auditory and cognitive processing tasks.

4.1. Auditory processing

The musicians had lower frequency discrimination thresholds than the non-musicians and this difference was statistically significant. Enhanced frequency discrimination performance of musicians has been documented previously (Micheyl et al., 2006; Spiegel and Watson, 1984). Similar to the current study, Micheyl et al. (2006) found that musicians with classical music training of more than 10 years had lower frequency discrimination thresholds than non-musicians. Kishon-Rabin et al. (2001) found that classical

[1] We also confirmed this using a non-parametric Spearman's correlation. Performance on aSL and frequency discrimination was not correlated for either group (musicians: $r_s = 0.35$, p = 0.17; non-musicians: $r_s = -0.4$, p = 0.11).

Table 4
Means, medians, SDs and effect sizes for performance on tests of cognition. Significant results are in bold font.

Test	Musicians			Non-musicians			Statistics		
	Mean	Median	SD	Mean	Median	SD	U	*p*	r
Forwards digit span scaled scores	10.9	11	1.8	9.5	10	2.8	110.5	0.2	−0.24
Backwards digit span scaled scores	**12.2**	**13**	**2.5**	**9.5**	**9.5**	**2.1**	**57**	**0.001**	**−0.53**
Stroop colour word interference score (ms)	134.3	110.8	80.7	154.5	104	163.9	145	0.8	−0.04
Sustained auditory attention quotient	108.8	110	18.3	103.4	110	30.1	147.5	0.9	−0.03
Sustained visual attention quotient	106.1	111	18.0	96.9	105.5	28.7	105	0.1	−0.27

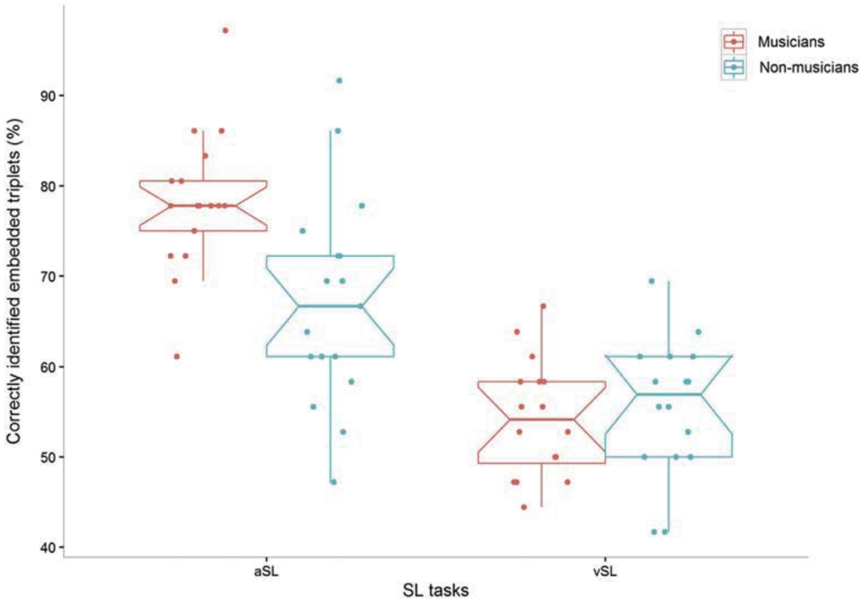

Fig. 3. Box and whisker plots showing the percentage of embedded triplets correctly identified by musicians and non-musicians on the aSL and vSL tasks. The box denotes 75th to 25th percentile (interquartile range). The upper whisker represents 75th percentile $+1.5 \times$ IQR and the lower whisker represents 25th percentile $-1.5 \times$ IQR. 50% indicates chance performance.

musicians had smaller frequency discrimination thresholds than contemporary (e.g. jazz, modern) musicians. It is noteworthy that the musicians in our study were also trained in classical music. Thus, smaller frequency discrimination thresholds may be due to emphasis on correct tuning during classical training (Micheyl et al., 2006).

We measured frequency discrimination at 1 kHz which has been suggested to be dominated by the use of temporal information that is, a phase locking mechanism (Moore, 2012). Musicians have been shown to have more precise temporal phase locking at the brainstem level as indexed by the frequency following response (FFR) (Lee et al., 2009). Better phase locking could be one of the reasons for finer frequency discrimination in musicians. Furthermore, it has been suggested that enhanced frequency discrimination could be due to better short-term memory representation or increased attention (Kishon-Rabin et al., 2001; Tervaniemi et al., 2005). However, for the musician group in our study, there was no correlation between frequency discrimination and performance on working memory and attention tasks. It is possible that the relationship between frequency discrimination and cognitive measures may be better probed by using different tests of attention and working memory than those used in the current study.

A right ear advantage (REA) was observed for both groups in the dichotic digits task. The difference in REA scores between the groups was not statistically significant. Musicians and non-musicians also had similar performance on the gap detection test. These findings are in agreement with previous literature comparing musicians and non-musicians on dichotic listening tasks involving digits (Nelson et al., 2003), and gaps in noise tasks (Ishii et al., 2006; Monteiro et al., 2010). Taken together, these findings suggest that superior performance of musicians may be limited to specific auditory tasks.

There were no group differences for detection of sinusoidal amplitude modulation of noise and discrimination of iterated rippled noise. Though a previous study (Lee et al., 2009) reported stronger encoding of temporal envelope cues in the brainstem of musicians using an electrophysiological technique (FFR), we found no evidence of this using behavioural measures. Enhancements in electrophysiological measures (such as increase in the amplitude of FFR) might not necessarily translate to enhancements in behavioural measures (Bidelman et al., 2011). Further research using more comparable behavioural and electrophysiological paradigms is needed to provide information about temporal envelope processing in musicians.

Musicians and non-musicians had similar speech perception in noise scores measured through the LiSN-S test. This is in line with three studies that also investigated speech perception in noise using a variety of tests (e.g. Boebinger et al., 2015; Fuller et al., 2014;

Ruggles et al., 2014). In contrast, in other studies of speech perception, musicians outperformed non-musicians (Parbery-Clark et al., 2009, 2011, 2012). These differences could be due to the different tests used across the studies. However, Ruggles et al. (2014), using the same tests and participant criteria as used by Parbery-Clark et al. (2009), found no significant advantage of being a musician on speech perception in noise tests (Experiment 2). Additionally, findings from Parbery-Clark et al. (2009) suggested that group differences may be observed only when the masking tasks are difficult, for example when the target and maskers are co-located. We administered two subtests of the LISN-S test, where the target speech and masker were co-located (different voices-0° and same voice-0°). However, there was no difference between the groups even when maskers were co-located and had no fundamental frequency cues (the hardest condition, i.e. same voice-0°). A recent study also found no significant group differences when the masker and target speech were co-located (Clayton et al., 2016). Further research involving stricter criteria for selection of participants (for example, recruitment of only professional musicians) and assessment of performance on a wider range of speech in noise tests may assist in understanding the relationship between speech perception in noise and musical experience. Overall, our results suggest that musical expertise might be associated with enhancements in only a subset of auditory perceptual skills.

4.2. Cognition

The current findings add to the converging evidence for musicians' superior performance on working memory tasks (Chan et al., 1998; Franklin et al., 2008; Ho et al., 2003). It has been suggested that musicians allocate more brain resources as the working memory load increases. This was evidenced by larger blood oxygenation-level dependent (BOLD) signal measured with functional magnetic resonance imaging (fMRI) in musicians compared to non-musicians during an n-back working memory task (Pallesen et al., 2010).

Interestingly, we found that musicians had significantly better scores on backward digit span but not on forward digit span. It has been suggested that distinct cognitive processes are tapped by forward and backward digit span tests and this may be why we observed that musicians outperformed non-musicians on backward but not forward digit span. For instance, the difference in performance on the forward and backward digit span tests may be explained by two theoretical approaches – the complexity view and the representational view (Rosen and Engle, 1997).

According to the complexity view, backward recall involves considerable attentional demands, with manipulation of digits held in short-term memory, and is thus considered to be a part of executive function processes tapping into working memory (Rosen and Engle, 1997). The second approach, the representational view, argues that backward recall may involve specific visuospatial processing where items are represented in a spatial array for easier reversal (Li and Lewandowsky, 1995; St Clair-Thompson and Allen, 2013). During the course of training and practice, musicians are required to learn melody and memorize sequences of notes either by learning through ear or through visual memory of music notations. A reasonable hypothesis to explain these group differences is that musicians improve their working memory storage through practice (consistent with the complexity view). Alternatively, musicians may employ visual strategies (such as visualizing each digit) to assist them with backward recall (consistent with the representational view). Anecdotally, a few musicians in our study reported using visual strategies during the backward recall task.

In the current study, musicians did not have enhanced performance on the visual Stroop task which is consistent with evidence that musical experience may be linked with specific components of executive function such as backwards recall but not inhibition (Boebinger et al., 2015; Clayton et al., 2016; Zuk et al., 2014). It is noteworthy that Bialystok and DePape (2009) using an *auditory* Stroop task demonstrated enhanced inhibition in musicians. Further studies are required to understand the influences of modality on executive function tasks.

Although it has been suggested that musical training contributes to enhanced auditory but not visual attention (Strait et al., 2010), our results indicated no enhancements in either visual or auditory sustained attention. This could be attributed to the different tasks and types of attention measured in the two studies. Strait and colleagues measured *alertness* in one modality at a time by comparing reaction times in the presence or absence of a variable delay cue. In contrast, our study used the IVA-CPT test (Table 2) which measured *sustained* attention in auditory and visual modalities concurrently by accounting for accuracy as well as reaction time in the presence of a distracting stimulus. Overall, the mixed pattern of results observed between this and other studies suggests that the relationship between musical expertise and general cognitive abilities is complex and needs further investigation.

4.3. Statistical learning

To the best of our knowledge, this is the first study showing behavioural differences in aSL between musicians and non-musicians using an embedded triplet task comprising tones. In previous studies that have assessed aSL using both electrophysiological and behavioural measures in musicians and non-musicians, group differences were observed only for electrophysiological measures (François and Schön, 2011; Paraskevopoulos et al., 2012). However, it should be noted that behavioural test results in the latter study did not differ significantly from chance for either group. In our study, both groups showed a mean aSL that was significantly better than chance, however, musicians outperformed non-musicians.

Though the current research design cannot address the question of whether musical training causes enhancement of aSL in adults, a randomized longitudinal training study of 8-year old children suggested a causal relationship between musical training and aSL (François et al., 2013). The enhanced aSL for musicians in our study may be attributed to the fact that Western tonal music is based on a strong system of regularities between musical events, such as notes and chords (Tillmann et al., 2000). Our findings suggest that long-term musical expertise could help form associations between successive stimuli and group them into distinctive units (triplets in our case) purely based on the transitional probabilities. Although the triplets we used did not follow the rules of any standard music composition, it could be argued that musicians' pre-existing knowledge regarding relative pitches of the Western scale could make it easier to detect statistical regularities amongst tones using the Western scale (Loui, 2012). Using a new, unfamiliar Bohlen-Pierce scale, Loui and colleagues reported no differences in performance of musically trained and untrained subjects for melody recognition and rule generalization (Loui et al., 2010). However, it should be noted that the musician group in that study had less musical training than the participants in our study (range of 5–14 years with an average of 9.6 years). In contrast, the musicians in our study had at least 10 years' experience of singing or playing music (range of 10–52 years with an average of 21.6 years). A systematic study using professional musicians or musicians with at least 10 years of musical training might help to clarify if the duration of the musical training is relevant in statistical learning of unfamiliar musical scales.

Notwithstanding the fact that some previous research has shown that musical expertise is associated with an enhanced visuospatial iconic representation (Gromko and Poorman, 1998), and visual imagery (Neuhoff et al., 2002), musicians and non-musicians in the current study had comparable performance on the vSL task. This finding could be taken as support for an experience-dependent modality specific enhancement of SL that is, the musicians' advantage in SL was only seen in one modality (aSL) and not the other modality (vSL). It is possible that more efficient usage of visuospatial perception and imagery in musicians (Brochard et al., 2004) might translate into enhancements in vSL measured using stimuli that contains spatial regularities. Further research is needed to explore these possibilities.

Better performance in the aSL task than in the vSL task by both groups suggests that SL proceeds differently across the visual and auditory modalities (Conway and Christiansen, 2005, 2009). Additionally, there was no significant correlation of the scores for the aSL and vSL tasks. This is in line with recent research (Frost et al., 2015; Siegelman and Frost, 2015). Given that musical practice involves the integration of multi-modal stimuli, future studies may further explore the role of musical expertise in SL using an auditory-visual SL task.

4.4. Relationship between statistical learning and measures of auditory processing and cognition

Shook et al. (2013) postulated that enhanced SL in musicians may be due to their enhanced temporal processing of auditory information as reflected in temporal discrimination, gap detection and rhythm perception tasks (Rammsayer and Altenmüller, 2006). However, in our study, performance on both SL tasks was independent of performance on any auditory processing task for both musicians and non-musicians. Future studies could use an aSL task involving stimuli with closer frequency separation to advance our knowledge about the relationship between aSL and frequency discrimination in musicians. In addition, further investigation using a variety of auditory processing tasks involving music perception (such as chord perception, melody discrimination) could help shed light on what strategies musicians use for extracting distributional cues.

Our finding that SL performance was not related to cognitive ability is in line with previous studies where SL was not associated with measures of verbal reasoning, intelligence and working memory (Kaufman et al., 2010). Siegelman and Frost (2015) also reported that there was no relationship between scores on SL (aSL and vSL tasks using the embedded triplet paradigm) and a large battery of cognitive tests (working memory, verbal working memory, rapid automatized naming, and switch task). These findings point towards the independence of SL from other general cognitive abilities. However, it should be noted that the relationship between SL and cognitive abilities such as working memory is complex and warrants further investigation. For instance, Arciuli and Simpson (2011) argued that implicit rather than explicit working memory tasks might reveal a relationship with SL. Furthermore, SL performance measured through 2AFC trials may involve additional processes such as decision making (François et al., 2014). Future studies may investigate the relationship between SL and cognitive abilities by incorporating concurrent cognitive tasks as SL takes place and using a wide battery of SL tasks (including SL tasks that do not rely on 2AFC trials).

It has been hypothesized that SL operates automatically with little or no dependence on executive attentional resources (Turk-Browne et al., 2005). Indeed, our results showed that SL performance was not correlated with measures of attention and inhibition. For both SL tasks used in our study, the participants were engaged in a cover task during familiarization that required them to process the stimuli in a different way (to detect oddballs rather than to extract statistical information per se). It is notable that the participants were successful in recognising embedded triplets in both the aSL and vSL tasks despite this competing demand that meant they were actively trying to process the stimuli in another way. As long as participants attend to the relevant stimuli, aSL and vSL may indeed operate automatically.

As music consists of a variety of temporal patterns, future research may investigate the link between aSL and other aspects of auditory processing, such as rhythm perception. Finally, a longitudinal study involving music training and measurement of SL, along with a comprehensive auditory processing test battery could shed some light on causality and the mechanisms underlying enhanced aSL.

4.5. Summary

In summary, using a set of auditory processing, cognitive, and statistical learning measures, we found that long-term musical expertise is associated with better performance on frequency discrimination and backward digit span tasks. We also present the first empirical evidence of better aSL, but not vSL for musicians than for non-musicians as evaluated using the embedded triplet paradigm. Performance on SL tasks was not correlated with performance on any other auditory processing or cognition measures.

Acknowledgements

We would like to thank the editor Brian Moore and three anonymous reviewers for their helpful comments. This work was funded by the Australia Awards Scholarships and the HEARing CRC, established and supported under the Cooperative Research Centres Program – Business Australia. This work was also supported by the Australian Research Council Centre of Excellence for Cognition and its Disorders (CE110001021). Demuth was supported by a Laureate Fellowship granted by the Australian Research Council (FL130100014). Arciuli was supported by a Future Fellowship granted by the Australian Research Council (FT130101570).

References

Abla, D., Katahira, K., Okanoya, K., 2008. On-line assessment of statistical learning by event-related potentials. J. Cognitive Neurosci. 20 (6), 952–964.
Abla, D., Okanoya, K., 2009. Visual statistical learning of shape sequences: an ERP study. Neurosci. Res. 64 (2), 185–190. http://dx.doi.org/10.1016/j.neures.2009.02.013.
Arciuli, J., Simpson, I.C., 2011. Statistical learning in typically developing children: the role of age and speed of stimulus presentation. Dev. Sci. 14 (3), 464–473. http://dx.doi.org/10.1111/j.1467-7687.2009.00937.x.
Arciuli, J., Simpson, I.C., 2012a. Statistical learning is lasting and consistent over time. Neurosci. Lett. 517 (2), 133–135. http://dx.doi.org/10.1016/j.neulet.2012.04.045.
Arciuli, J., Simpson, I.C., 2012b. Statistical learning is related to reading ability in children and adults. Cognitive Sci. 36 (2), 286–304. http://dx.doi.org/10.1111/j.1551-6709.2011.01200.x.
Arciuli, J., von Koss Torkildsen, J., 2012. Advancing our understanding of the link between statistical learning and language acquisition: the need for longitudinal data. Front. Psychol. 3, 324. http://dx.doi.org/10.3389/fpsyg.2012.00324.
Baker, R.J., Jayewardene, D., Sayle, C., Saeed, S., 2008. Failure to find asymmetry in auditory gap detection. Laterality 13 (1), 1–21.
Bialystok, E., DePape, A.-M., 2009. Musical expertise, bilingualism, and executive functioning. J. Exp. Psychol. Hum. Percept. Perform. 35 (2), 565–574.
Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cognition 77 (1), 1–10.
Boebinger, D., Evans, S., Rosen, S., Lima, C.F., Manly, T., Scott, S.K., 2015. Musicians and non-musicians are equally adept at perceiving masked speech. J. Acoust. Soc. Am. 137 (1), 378–387. http://dx.doi.org/10.1121/1.4904537.
Brandler, S., Rammsayer, T.H., 2003. Differences in mental abilities between musicians and non-musicians. Psychol. Music 31 (2), 123–138.
Brochard, R., Dufour, A., Després, O., 2004. Effect of musical expertise on visuospatial abilities: evidence from reaction times and mental imagery. Brain

Cognition 54 (2), 103–109. http://dx.doi.org/10.1016/S0278-2626(03)00264-1.
Cameron, S., Dillon, H., 2007. Development of the listening in spatialized noise-sentences test (LISN-S). Ear Hear. 28 (2), 196–211.
Chan, A.S., Ho, Y.-C., Cheung, M.-C., 1998. Music training improves verbal memory. Nature 396 (6707), 128–128.
Clayton, K.K., Swaminathan, J., Yazdanbakhsh, A., Zuk, J., Patel, A.D., Kidd Jr., G., 2016. Executive function, visual attention and the cocktail party problem in musicians and non-musicians. PloS one 11 (7), e0157638. http://dx.doi.org/10.1371/journal.pone.0157638.
Conway, C.M., Bauernschmidt, A., Huang, S.S., Pisoni, D.B., 2010. Implicit statistical learning in language processing: word predictability is the key. Cognition Int. J. Cognitive Sci. 114 (3), 356–371. http://dx.doi.org/10.1016/j.cognition.2009.10.009.
Conway, C.M., Christiansen, M.H., 2005. Modality-constrained statistical learning of tactile, visual, and auditory sequences. J. Exp. Psychol. Learn. Mem. Cognition 31 (1), 24.
Conway, C.M., Christiansen, M.H., 2009. Seeing and hearing in space and time: effects of modality and presentation rate on implicit statistical learning. Eur. J. Cognitive Psychol. 21 (4), 561–580.
Fine, P.A., Moore, B.C., 1993. Frequency analysis and musical ability. Music Percept. An Interdiscip. J. 11 (1), 39–53.
François, C., Chobert, J., Besson, M., Schön, D., 2013. Music training for the development of speech segmentation. Cereb. Cortex 23 (9), 2038–2043.
François, C., Jaillet, F., Takerkart, S., Schön, D., 2014. Faster sound stream segmentation in musicians than in nonmusicians. PloS one 9 (7), e101340.
François, C., Schön, D., 2010. Learning of musical and linguistic structures: comparing event-related potentials and behavior. Neuroreport 21 (14), 928–932. http://dx.doi.org/10.1097/WNR.0b013e32833ddd5e.
François, C., Schön, D., 2011. Musical expertise boosts implicit learning of both musical and linguistic structures. Cereb. Cortex 21 (10), 2357–2365. http://dx.doi.org/10.1093/cercor/bhr022.
François, C., Schön, D., 2014. Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: the role of musical practice. Hear. Res. 308, 122–128.
Franklin, M.S., Moore, K.S., Yip, C.-Y., Jonides, J., Rattray, K., Moher, J., 2008. The effects of musical training on verbal memory. Psychol. Music 36 (3), 353–365.
Frost, R., Armstrong, B.C., Siegelman, N., Christiansen, M.H., 2015. Domain generality versus modality specificity: the paradox of statistical learning. Trends Cognitive Sci. 19 (3), 117–125.
Fuller, C.D., Galvin, J.J., Maat, B., Free, R.H., Başkent, D., 2014. The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? Front. Neurosci. 8, 179. http://dx.doi.org/10.3389/fnins.2014.00179.
Gaser, C., Schlaug, G., 2003. Brain structures differ between musicians and non-musicians. J. Neurosci. 23 (27), 9240–9245.
Gromko, J.E., Poorman, A.S., 1998. Does perceptual-motor performance enhance perception of patterned art music? Music. Sci. 2 (2), 157–170.
Helmbold, N., Rammsayer, T., Altenmüller, E., 2005. Differences in primary mental abilities between musicians and nonmusicians. J. Individ. Differ. 26 (2), 74–85.
Ho, Y.-C., Cheung, M.-C., Chan, A.S., 2003. Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. Neuropsychology 17 (3), 439–450.
Iliadou, V.V., Bamiou, D.E., Chermak, G.D., Nimatoudis, I., 2014. Comparison of two tests of auditory temporal resolution in children with central auditory processing disorder, adults with psychosis, and adult professional musicians. Int. J. Audiol. 53 (8), 507–513. http://dx.doi.org/10.3109/14992027.2014.900576.
Ishii, C., Arashiro, P.M., Pereira, L.D., 2006. Ordering and temporal resolution in professional singers and in well tuned and out of tune amateur singers. Pró-Fono Rev. Atualização Científica 18 (3), 285–292.
Jakobson, L.S., Lewycky, S.T., Kilgour, A.R., Stoesz, B.M., 2008. Memory for verbal and visual material in highly trained musicians. Music Percept. An Interdiscip. J. 26 (1), 41–55.
Kaufman, S.B., DeYoung, C.G., Gray, J.R., Jiménez, L., Brown, J., Mackintosh, N., 2010. Implicit learning as an ability. Cognition Int. J. Cognitive Sci. 116 (3), 321–340. http://dx.doi.org/10.1016/j.cognition.2010.05.011.
Kishon-Rabin, L., Amir, O., Vexler, Y., Zaltz, Y., 2001. Pitch discrimination: are professional musicians better than non-musicians? J. Basic Clin. Physiol. Pharmacol. 12 (2), 125–144.
Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective subcortical enhancement of musical intervals in musicians. J. Neurosci. 29 (18), 5832–5840.
Lee, Y.-s., Lu, M.-j, Ko, H.-p., 2007. Effects of skill training on working memory capacity. Learn. Instr. 17 (3), 336–344.
Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49 (2B), 467–477.
Li, S.-C., Lewandowsky, S., 1995. Forward and backward recall: different retrieval processes. J. Exp. Psychol. Learn. Mem. Cognition 21 (4), 837.
Loui, P., 2012. Statistical learning – What can music tell us?us. In: Rebuschat, P., Williams, J. (Eds.), Statistical Learning and Language Acquisition. Mouton de Gruyter, Berlin, Germany, pp. 433–462.
Loui, P., Wessel, D.L., Kam, C.L.H., 2010. Humans rapidly learn grammatical structure in a new musical scale. Music Percept. An Interdiscip. J. 27 (5), 377–388.
Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A.J., 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hear. Res. 219 (1), 36–47.
Monteiro, R.A.M., Nascimento, F.M., Soares, C.D., Ferreira, M.I.D.d.C., 2010. Temporal resolution abilities in musicians and no musicians violinists. Arq. Int. Otorrinolaringol. 14 (3), 302–308.
Moore, B.C.J., 2012. An Introduction to the Psychology of Hearing. Brill, Netherlands.
Münte, T.F., Altenmüller, E., Jäncke, L., 2002. The musician's brain as a model of neuroplasticity. Nat. Rev. Neurosci. 3 (6), 473–478.
Nelson, D.M., Wilson, R.H., Kornhass, S., 2003. Performance of musicians and nonmusicians on dichotic chords, dichotic CVs, and dichotic digits. J. Am. Acad. Audiol. 14 (10), 536–544.
Neuhoff, J.G., Knight, R., Wayand, J., 2002. Pitch change, sonification, and musical expertise: Which way is up?. In: Paper Presented at the International Conference on Auditory Display.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113.
Pallesen, K.J., Brattico, E., Bailey, C.J., Korvenoja, A., Koivisto, J., Gjedde, A., Carlson, S., 2010. Cognitive control in auditory working memory is enhanced in musicians. PloS one 5 (6), e11120.
Paraskevopoulos, E., Kuchenbuch, A., Herholz, S.C., Pantev, C., 2012. Statistical learning effects in musicians and non-musicians: an MEG study. Neuropsychologia 50 (2), 341–349. http://dx.doi.org/10.1016/j.neuropsychologia.2011.12.007.
Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009. Musician enhancement for speech-in-noise. Ear Hear. 30 (6), 653–661.
Parbery-Clark, A., Strait, D.L., Kraus, N., 2011. Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. Neuropsychologia 49 (12), 3338–3345.
Parbery-Clark, A., Tierney, A., Strait, D.L., Kraus, N., 2012. Musicians have fine-tuned neural distinction of speech syllables. Neuroscience 219, 111–119.
Patel, A.D., 2010. Music, Language, and the Brain. Oxford university press.
Patel, A.D., 2011. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Front. Psychol. 2, 142. http://dx.doi.org/10.3389/fpsyg.2011.00142.
Peter, V., Wong, K., Narne, V.K., Sharma, M., Purdy, S.C., McMahon, C., 2014. Assessing spectral and temporal processing in children and adults using temporal modulation transfer function (TMTF), iterated ripple noise (IRN) perception, and spectral ripple discrimination (SRD). J. Am. Acad. Audiol. 25 (2), 1–9.
Rammsayer, T., Altenmüller, E., 2006. Temporal information processing in musicians and nonmusicians. Music Percept. An Interdiscip. J. 24 (1), 37–48.
Rodrigues, A.C., Guerra, L.B., Loureiro, M.A., 2007. Visual attention in musicians and non-musicians: a comparative study. In: Paper Presented at the Proceedings of the 3rd International Conference on Interdisciplinary Musicology.
Rohrmeier, M., Rebuschat, P., 2012. Implicit learning and acquisition of music. Top. Cognitive Sci. 4 (4), 525–553.
Rosen, V.M., Engle, R.W., 1997. Forward and backward serial recall. Intelligence 25 (1), 37–47.
Ruggles, D.R., Freyman, R.L., Oxenham, A.J., 2014. Influence of musical training on understanding voiced and whispered speech in noise. PloS one 9 (1), e86980.
Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-Month-old infants. Science 274 (5294), 1926–1928.
Saffran, J.R., Johnson, E.K., Aslin, R.N., Newport, E.L., 1999. Statistical learning of tone sequences by human infants and adults. Cognition Int. J. Cognitive Sci. 70 (1), 27–52.
Schlaug, G., Norton, A., Overy, K., Winner, E., 2005. Effects of music training on the child's brain and cognitive development. Ann. N. Y. Acad. Sci. 1060 (1), 219–230.
Schön, D., François, C., 2011. Musical expertise and statistical learning of musical and linguistic structures. Front. Psychol. 2, 167. http://dx.doi.org/10.3389/fpsyg.2011.00167.
Semel, E.M., Wiig, E.H., Secord, W., 2006. CELF 4: Clinical Evaluation of Language Fundamentals. Pearson: Psychological Corporation.
Shook, A., Marian, V., Bartolotti, J., Schroeder, S.R., 2013. Musical experience influences statistical learning of a novel language. Am. J. Psychol. 126 (1), 95.
Siegelman, N., Frost, R., 2015. Statistical learning as an individual ability: theoretical perspectives and empirical evidence. J. Mem. Lang. 81, 105–120.
Skoe, E., Krizman, J., Spitzer, E., Kraus, N., 2013. The auditory brainstem is a barometer of rapid auditory learning. Neuroscience 243, 104–114. http://dx.doi.org/10.1016/j.neuroscience.2013.03.009.
Špajdel, M., Jariabková, K., Riečanský, I., 2007. The influence of musical experience on lateralisation of auditory processing. Laterality 12 (6), 487–499.
Spiegel, M.F., Watson, C.S., 1984. Performance on frequency-discrimination tasks by musicians and nonmusicians. J. Acoust. Soc. Am. 76 (6), 1690–1695.
St Clair-Thompson, H.L., Allen, R.J., 2013. Are forward and backward recall the same? A dual-task study of digit recall. Mem. Cognition 41 (4), 519–532.
Stevens, D.J., Arciuli, J., Anderson, D.I., 2015. Concurrent movement impairs incidental but not intentional statistical learning. Cognitive Sci. 39 (5), 1081–1098.
Strait, D.L., Kraus, N., 2011. Playing music for a smarter ear: cognitive, perceptual and neurobiological evidence. Music Percept. An Interdiscip. J. 29 (2), 133.
Strait, D.L., Kraus, N., Parbery-Clark, A., Ashley, R., 2010. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. Hear. Res. 261 (1), 22–29.
Strouse, A., Wilson, R.H., 1999. Recognition of one-, two-, and three-pair dichotic digits under free and directed recall. J. Am. Acad. Audiol. 10 (10), 557–571.
Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., Schröger, E., 2005. Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. Exp. Brain Res. 161 (1), 1–10.
Tillmann, B., Bharucha, J.J., Bigand, E., 2000. Implicit learning of tonality: a self-organizing approach. Psychol. Rev. 107 (4), 885–913.
Turk-Browne, N.B., Junge, J., Scholl, B.J., 2005. The automaticity of visual statistical

learning. J. Exp. Psychol. General 134 (4), 552–564. http://dx.doi.org/10.1037/0096-3445.134.4.552.
Turner, A., Sandford, J., 1995. A normative study of IVA: Integrated visual and auditory continuous performance test. In: Paper Presented at the Annual Convention of the American Psychological Association.
Van der Elst, W., Van Boxtel, M.P., Van Breukelen, G.J., Jolles, J., 2006. The Stroop color-word test: influence of age, sex, and education; and normative data for a large sample across the adult age range. Assessment 13 (1), 62–79. http://dx.doi.org/10.1177/1073191105283427.
Yost, W.A., 1996. Pitch strength of iterated rippled noise. J. Acoust. Soc. Am. 100 (5), 3329–3335.
Zendel, B.R., Alain, C., 2012. Musicians experience less age-related decline in central auditory processing. Psychol. Aging 27 (2), 410. http://dx.doi.org/10.1037/a0024816.
Zuk, J., Benjamin, C., Kenyon, A., Gaab, N., 2014. Behavioral and neural correlates of executive functioning in musicians and non-musicians. PloS one 9 (6), e99868. http://dx.doi.org/10.1371/journal.pone.0099868.

THE END