The influence of music experience on nonnative phonological perception

Lim, Samantha Jieying

2018

Lim, S. J. (2018). The influence of music experience on nonnative phonological perception.
Master's thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/74594>

<https://doi.org/10.32657/10356/74594>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**THE INFLUENCE OF MUSIC EXPERIENCE
ON NONNATIVE PHONOLOGICAL
PERCEPTION**

SAMANTHA LIM JIE YING
SCHOOL OF HUMANITIES
2018

The Influence of Music Experience on Nonnative Phonological Perception

Samantha Lim JieYing

School of Humanities

A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for the
degree of Master of Arts

2018

The Influence of Music Experience on
Nonnative Phonological Perception

CONTENTS

ACKNOWLEDGMENTS	vii
LIST OF TABLES AND FIGURES	viii
SYNOPSIS	1
ABSTRACT	2
1 INTRODUCTION	3
1.1 Motivation	3
1.2 Literature Background	4
1.2.1 Broad Parallels in Music and Language	4
1.2.2 Acoustic Similarities	5
1.3 Acoustic Transfers of Music to Language	5
1.3.1 Pitch	5
1.3.2 Timbre	8
1.3.3 Duration	8
1.3.4 Summary	12
1.4 Theoretical Frameworks on How Music Influences Language	12
1.4.1 Top-Down and Bottom-Up	12
1.4.2 Statistical Learning Models	15
1.5 Research Gap	16
1.6 The Study	17
1.7 Research Questions	17
2 METHODS AND STATISTICAL ANALYSIS	20
2.1 Exploratory Study	20
2.1.1 Voicing Categories	20
2.1.1a Voicing Categories in Hindi	20
2.1.1b Voicing Categories in English	21
2.1.1c Voicing Categories in Mandarin Chinese and Malay	21
2.1.1d Voicing Categories in Singapore Language Varieties	22
2.1.2 Participants	23
2.1.3 Methods	23
2.1.3a Materials	23
2.1.3b Task	24
2.1.3c Measurement	24
2.1.4 Results	25
2.1.5 Summary	26
2.2 Experiment 1	27
2.2.1 Participants	27
2.2.1a Participant Exclusion	28
2.2.2 Methods	29
2.2.2a Materials	29

2.2.2b	Tasks	30
2.2.3	Results	34
2.2.3a	AX Discrimination Test	34
2.2.3b	Ordered Discrimination Test	36
2.3	Experiment 2	42
2.3.1	Participants	42
2.3.2	Methods	43
2.3.2a	Materials	43
2.3.2b	Tasks	43
2.3.3	Results	44
2.3.3a	AX Discrimination	44
2.2.3b	Categorization	46
3	DISCUSSION	48
3.1	Music Experience as a Template for Processing Sounds	48
3.2	Music Experience as a Facilitator in Language Learning	49
3.3	The Role of Other Components in Music Experience	49
3.4	Consequences for Theoretical Frameworks	50
3.5	Future Directions	52
3.6	Conclusion	53
4	REFERENCES	54
5	APPENDICES	72
5.1	Appendix A	72
5.2	Appendix B	76
5.3	Appendix C	77

Acknowledgements

This thesis could not have appeared without the assistance of *many* wonderful people. It might be impossible to list each name on this small page, but I hope to be able to formally thank some and convey to all my sincere gratitude for their help and thoughts.

First, I would like very much to thank my supervisor Dr. Francis Wong Chun Kit. His academic guidance instilled and reinforced the values of discipline, resilience, and discretion curiously reminiscent of a triple package tiger parents purportedly present their offspring. He also taught me that, when thrown into a fire of “you can do this,” one may not necessarily have to sink or swim: an alternative is to float, and perhaps with time, swimmingly.

Next, special thanks and acknowledgement to Prof Alice Chan for her kindness, generous support and interest in my research project. She always ensured we were well-equipped, and I am truly grateful.

I also wish to extend my deepest gratitude and love to my lab friends. Their friendships enriched my life with fun, laughter and lab drama. I owe Lau Fun, our super intelligent lab socialite, a gigantic heap of thank yous for her gracious help and invaluable advice, from posing mock questions profs might ask to being there as a good friend when times got challenging. Special thanks to Kastoori, who improved my education greatly in other ways. Now, I will never forget how imperative it is to perform regressions and correlations, and that /kü-kü/ actually refers to human beings and not birds. To the rest of the gang –Sophia, Galston, Joey and Yvonne, for patiently listening to an entire year of presentations popped fortnightly. And, not least, Firqin, who helped me better understand that stress is best taken in bite sizes. You all are the very best!

Of course, none of this would also be possible without my participants. Each one of them was an integral part of this project, and I owe them my gratitude.

I must also thank my dear friends and family for their care and concern, especially my mom and dad, who put up with late nights home, stretched the internet curfew, and constantly reminded me of their love. Thank you!

Finally, I thank God, the Father of lights, who gives what is good and perfect.

List of Tables and Figures

Figure 1. Schematic of release time and voice onset time	10
Figure 2. Depiction of aspiration in a speech waveform	11
Table 1. Hindi Stop Contrasts in Monosyllabic Tokens	21
Table 2. Voicing Categories in Malay, English, Mandarin Chinese, and Hindi	22
Table 3. Participant Language Background Information	23
Figure 3. Word list of minimally contrastive stop tokens in pilot study	24
Figure 4. Example of measurement method for VOT values	24
Table 4. Production of Voiced and Voiceless Stops by Native Singapore English Speakers	25
Table 5. Summary of Voicing Contrasts in Singapore Language Varieties Discussed by Ng, 2005 ...	26
Table 6. Summary of Group Statistics	28
Table 7. Voicing Contrast Examples	30
Figure 5. Mean discrimination scores for VOT test contrasts across groups	34
Figure 6. Musician and nonmusician average discrimination scores across voicing contrast sets	35
Figure 7. Average categorization scores across musician and nonmusician groups	37
Figure 8. Mean categorization scores of musicians and nonmusicians	38
Figure 9. Average correct responses in ordered discrimination of all VOT contrasts across musician and nonmusician groups	39
Figure 10. Ordered discrimination scores across contrast sets	39
Figure 11. Group differences across keyboard, string, and wind musicians in ordered discrimination of VOT contrasts	41
Figure 12. Mean ordered discrimination score as a function of music training and perception priming condition	42
Figure 13. Schematic of categorization task	44
Figure 14. Mean total score in discriminating dental-retroflex contrasts across musicians and nonmusicians	45
Figure 15. Mean categorization score across groups	46

Music and speech are two highly complex systems with shared spectral and temporal acoustic features that involve similar cognitive processes. Hence, it is not surprising that correlations have been found between music experience and linguistic processing. Previous research has largely demonstrated music transfer advantages for perception of prosodic speech features and theoretical frameworks posited to study how music experience can contribute to speech processing. Building on past work, this thesis concerns positive transfers of music experience to segmental feature processing that extend beyond the native domain –a topic not much explored. It examines how perceptual accuracy of phonetic contrasts with minute timing and timbre differences may be influenced by one’s music experience. It further explores the effect of different components of music experience on perception, namely music aptitude, training, exposure, and also instrument specialization.

ABSTRACT

Music experience has been found to influence language processing. Previous studies reveal differences in musicians and nonmusicians across a range of speech processing tasks. A majority of these studies examine music-related transfer benefits to processing prosodic speech segments that contain shared acoustic properties in music. However, little is known about positive transfers to segmental speech processing as they are far transfers, and not immediately evident. Only a few studies have examined the transfers to segmental processing and even fewer explore such transfers beyond the native domain. Thus, this study seeks to evaluate in greater detail possible music-related transfer advantages for processing nonnative segmental contrasts, specifically in voice onset time (VOT) and place of articulation (POA), i.e. dental versus retroflex contrasts. It takes a different angle from previous studies by analyzing music experience in relation to aptitude, training, exposure and specialization and asks how these components in turn could affect transfer benefits to the perception of the target contrasts. AX, ordered discrimination and categorization tests were created containing phonotactically-acceptable Hindi nonce syllables that differed in VOT and POA. They assessed the perceptual accuracy of bilingual English-Chinese and English-Malay musicians and nonmusicians. Consistent with expectations, music experience resulted in positive transfers with significant differences reflected across groups. Musicians showed greater efficacy in perceiving and categorizing voicing and place of articulation contrasts. Results also suggest that positive transfers from music to speech are not merely dependent on musical training. Less investigated components of music experience, i.e. aptitude, sophistication, and music instrument specialization were found to predict successful performance across tests. Importantly, music exposure was not shown to affect the processing of nonnative contrasts. The significance of these findings will be discussed in relation to the widely-held belief that music training is the only contributor to music-related transfer benefits.

INTRODUCTION

1.1 Motivation

This research addresses the interplay of music experience and language processing. Music experience is decomposed into several factors to explore how each may drive positive transfers in speech. Linguistic theories are taken in account to discuss how positive transfers to speech perception happen. Two experiments were conducted to shed light on the way segmental speech in a nonnative language can be impacted by music experience.

In recent years, interest in the interaction of music and language has been evidenced by numerous comparative research studies across the two domains. The scope of comparison between them encompasses multidisciplinary studies that examine a wide range of speech processing tasks. The main questions asked in all these studies can be summed to three: 1) if there are shared cognitive and perceptual processing mechanisms apropos of observable transfers from one domain to another; 2) where the processing is represented; and 3) how positive transfers are possible across different domains. In linguistics, past studies largely analyze these questions by way of correlational and psycholinguistic methods, which can be applied to our study. We will examine these questions of interest in greater depth.

The first question relates to the existence of transfer benefits in processing mechanisms across music and speech domains. For this we focus on transfer benefits in auditory processing. Given that speech and music are composed by sequences of sound units in time, fine-tuning one's ability to perceive these sound units in an acoustically demanding domain such as music will likely lead to a fine-tuned perception of similar units in speech. Within auditory processing, positive transfers happen across various aspects, namely pitch, duration, and timbre. This study will explore positive transfers to perception of duration in segmental speech. This differs from many other studies which examine duration in the context of prosodic speech features, e.g. stress, meter, with larger and more salient duration differences than those in phonetic contrasts. Given that music experience adds to one's exposure to rhythmic and timing features, we wonder if it could affect one's sensitivity to smaller duration contrasts in nonnative segmental speech.

The second question concerns processing mechanisms responsible for the transfer benefits. There are many levels at which one may choose to do the examination, including those from phonological and neurophysiological angles. While our study does not delve into this question, we will briefly

discuss frameworks proposed in the literature to explain the effect of music experience on underlying cognitive mechanisms, in relation to the third question often asked, that is, how positive transfers may occur.

To address the third question, relevant existing theories are discussed in relation to cross-domain auditory processing to explain positive transfers. In the next section, we will focus on past studies –what they report on acoustic similarities and positive transfers across music and language processing. After this, we will look at theoretical frameworks and identify the gap our current study addresses and why it is worthwhile to consider.

1.2 Literature Background

1.2.1 *Broad Parallels in Music and Language*

At first glance, music and language, two highly complex systems, appear to be highly dissociable. Yet it has been observed that from the standpoint of a naïve listener, both systems are similar expressions, used as a form of acoustic communication of sorts (Conference on Music, Language and the Brain, celebrating 25th anniversary of Lerdahl and Jackendoff’s Generative Theory of Tonal Music, 2008). Both consist of a combination of shared properties ordered into distinct perceptual categories.

Interest in the relationship between music and language has been long-standing. One of the earliest documented music-language inference dates back to 1816 where the mathematician and philosopher René Descartes noted resemblances across melody and language in his work “L’abrégé de musique,” decomposing music perception into three fundamental components –auditory, sensory, and aesthetic (Settari, 1997; Thomas & American Council of Learned Societies 1995).

Today this link continues to be of interest. A few have ventured to suggest evolutionary parallels between music and speech domains (Masataka, 2009; Mithen, 2006; Brown, 2001). Empirically, functional music-language associations investigated in behavioral and brain imaging studies show similarities neurophysically (Elmer, 2016; Peretz, Vuvan, Lagrois & Armony, 2015; Zatorre, 2013; Shahin, 2011; Bermudez, Lerch, Evans & Zatorre, 2009), structurally (Kunert, Willems, Casasanto, Patel & Hagoort, 2015; Peretz, Vuvan, Lagrois & Armony, 2015; Slevc & Okada, 2015; Herdener et al., 2014; Fedorenko, Patel, Casasanto, Winawer & Gibson, 2009; Hara et al., 2009; Koelsch, Gunter, Wittfoth & Sammler, 2005; Levitin & Menon, 2003; Patel, Gibson, Ratner, Besson & Holcomb, 1998); and most salient of all, acoustically (Perrachione, Fedorenko, Vinke, Gibson &

Dilley, 2013; Bidelman, Gandour & Krishnan, 2011; Musacchia, Sams, Skoe & Kraus, 2007; Magne, Schön & Besson, 2006; Schön, Magne & Besson, 2004).

1.2.2 Acoustic Similarities

The most arresting similarity across music and language domains concerns acoustic properties as both systems cannot exist without the decoding and parsing of auditory sequences. In fact, the anatomy of music and language may be defined by three basic features, namely pitch, timbre, and duration. These fundamental properties are processed by the same physical organs and have been shown to activate several similar neural processing modules (Abrams, Bhatara, Ryali, 2011; Zatorre & Schönwiesner, 2011; Koelsch et al., 2004; Levitin & Menon, 2003) even while both domains are encoded in functionally distinct ways.

It is then no surprise to find evidence of beneficial interactions between music and language processing. A large number of studies have found that music training may lead to perceptual advantages in speech processing tasks. Musicians were found to be highly sensitive to tonal (Marie, Delogu, Lampis, Belardinelli & Besson, 2011; Chandrasekaran, Krishnan & Gandour, 2009; Parbery-Clark, Skoe & Kraus, 2009) and speech-in-noise stimuli (Jain, Mohamed & Kumar, 2015; Tervaniemi, Just, Koelsch, Widmann & Schröger, 2005). In addition, music experience was shown to be closely related to better vowel classification (Reinke, He, Wang & Alain, 2003), phonemic/phonological awareness (Dege & Schwarzer, 2011; Gromko, 2005), phonological processing (Jones, Lucker, Zalewski, Brewer & Drayna, 2009; Anvari, Trainor, Woodside & Levy, 2002), pronunciation skill (Milovanov, Huotilainen, Valimaki, Esquef & Tervaniemi, 2008), encoding of speech cues (Strait, Parbery-Clark, Hittner & Kraus, 2012) and general linguistic ability (Shook, Marian, Bartolotti & Schroeder, 2013; Moreno et al., 2009; Slevc & Miyake, 2006).

For these earlier studies, music experience positively transfers to processing a diverse range of language tasks, yet the intrinsic elements that motivate this advantage were not clear. A recent wave of studies then focused on a more direct comparison of acoustic properties between domains, evaluating shared pitch, timbre, and duration features in music-to-speech transfers, which we will discuss below.

1.3 Acoustic Transfers of Music to Language

1.3.1 Pitch

One of the more-studied topic in music and language studies is pitch processing. Pitch is a perceptual unit that constitutes melodic sequences in music and intonational aspects in speech. In

pure tones, pitch is acoustically specified as the frequency or number of cycles a second in a sound wave (Krumhansl, 2000), while in complex sounds such as speech, it corresponds primarily to the fundamental frequency (F0), the lowest resonant frequency in a periodic wave (Davenport & Hannahs, 2005, pp. 61).

The ability to distinguish pitch is crucial to understanding music. In formal music training, musicians are frequently exposed to well-defined musical context and learn to distinguish a wide range of musical pitch. Given this, early experiments confirmed the intuition that music training increases sensitivity to pitch (Lee, Lekich & Zhang, 2014; Micheyl, Delhommeau, Perrot & Oxenham, 2006; Tervaniemi, Just, Koelsch, Widmann & Schröger, 2005; Kishon-Rabin, Amir, Vexler & Zaltz, 2001; Spiegel & Watson, 1984). Musicians have been found to show heightened sensitivity to pitch changes in music where music trained adults and children were better able to discriminate pitch incongruence at graduated levels than nonmusicians (Magne, Schön & Besson, 2006). Building upon these conclusions, other studies explored the idea of cross-domain transfers from musical to linguistic pitch processing.

In language, pitch is used to denote meaning at both phrasal and lexical levels. Phrasal pitch is commonly known as intonation, where the rise and fall of sounds in a phrase or sentence convey linguistically relevant information to the listener. Changes in F0 often contribute to modulation of phrasal pitch and also indicate affective states (Juslin & Laukker, 2001) and speaker identity relative to linguistic background (Barcroft & Sommers, 2014). In English, pitch changes at sentence level, or intonation, may be used for encoding pragmatic information. For example, a falling pitch at the end of an utterance “Darwin said he was going to the mall” would denote the finality of a stated fact in contrast to a rising pitch at the same position which might indicate doubt, a request for confirmation.

Studies have reported a musician advantage for detecting changes in both phrasal pitch and sentence intonation. There is evidence of enhanced brain activation response to specific affective prosody in phrases. Musicians were found to demonstrate marked neural differences compared to nonmusicians, notably in response to sad prosodic speech cues (Park et al., 2015). They also showed better accuracy in identifying a range of other emotions signaled by pitch variation in speech, including surprise, neutrality, happiness, fear, sadness, anger, disgust (Lima & Castro, 2011; Thompson, Schellenberg & Husain, 2004). Findings suggest improved emotion processing by musicians as a result of a great efficiency to perceive basic acoustic cues. Musicians also outperformed nonmusicians in detecting subtle pitch changes in sentence intonation in native (Deguchi et al., 2012) and nonnative languages (Deguchi et al., 2012; Dankovicová, House, Crooks

& Jones, 2007; Marques, Moreno, Castro & Besson, 2007). Moreover, musicians showed enhanced neural speech encoding compared to nonmusicians (Song, Skoe, Banai & Kraus, 2012; Kraus, Skoe, Parbery-Clark & Ashley, 2009). These results indicate a strong correlation between music training and a heightened awareness of pitch nuances in utterances.

In addition, pitch may be used to mark meaning change lexically. This is most typical in the context of tone languages. Tone languages, which comprise over 60% of the world's languages (Downing & Rialland, 2017; Bao, 2003 & Yip, 2002), employ pitch as a semantic feature of spoken word. The same word given different pitches or lexical tones denote different meanings. For instance, in Thai the syllable /mâi/ given a falling pitch means to “no; not” while a low pitch /màì/ means “new” and a rising pitch in /mǎi/ denotes an interrogative particle. In contrast to F0 in melodic pitch, meaningfully contrastive pitch in words may be perceptually categorized across several acoustic dimensions in relation to F0, namely height, contour and direction. Step changes from the onset to offset of the sound determine the relative pitch height of lexical tone, while rapid changes in F0 and formants may be used to reflect pitch contour and direction (Wang, Wang & Chen, 2013; Chandrasekaran, Gandour & Krishnan, 2007). These dimensions are important in aiding accurate discrimination of lexical tones.

In the widely-examined sphere of tone languages, a large number of studies have discovered that nonnative tone contrasts are better discriminated by musicians than nonmusicians. Music training was found to result in faster lexical tone discrimination and reflect increased mismatch negativity (MMN) amplitudes in brain response to differences in lexical tone differences (Tang, Xiong, Zhang, Dong & Nan, 2016). Adult musicians (Kühnis, Elmer, Meyer & Jäncke, 2013) and music trained children (Chobert, Marie, Francois, Schon & Besson, 2011) also demonstrated enhanced auditory sensitivity to duration changes in vowels and voice onset time and greater brain response when temporal differences were detected. For pitch discrimination, musician showed faster behavioral response and superior accuracy than nonmusicians in distinguishing pitch in contexts including violin sounds, low-pass filtered speech, and nonnative lexical Thai tones (Burnham, Brooker & Reid, 2015). Moreover, significant differences were reported where musician English native speakers were able to identify (Lee & Hung, 2008), discriminate and produce (Gottfried, Staby & Ziemer, 2004) nonnative tone contrasts in Mandarin with greater accuracy than nonmusicians

Aside from training, musical aptitude was found to positively correlate with sensitivity to tonal contrast in another language. Significant correlations between music aptitude scores, e.g. Advanced Measures of Musical Audiation (AMMA) and accuracy in discriminating nonnative Norwegian tonal contrasts were observed by Kempe, Bublitz & Brooks (2015). Likewise Delogu, Lampis, &

Olivetti Belardinelli (2006) reported that participants who demonstrated superior melodic ability in the Wing Musical Aptitude Test outperformed others in discriminating tonal changes in Mandarin tokens. The past studies indicate music-to-speech benefits which will be discussed in greater detail in relation to proposed frameworks motivating such transfers.

1.3.2 Timbre

Music experience have also resulted in positive transfers to processing spectral and harmonic speech features. Spectro-harmonic changes in time make up complex sound signals, and the combination of a spectrum (amplitude differences of individual harmonic frequencies) and amplitude envelope (the energy contour of a periodic wave) is often referred to as sound timbre. In music, timbre sets two sounds with the same pitch apart (Hass, 2013; Koelsch, 2011) differentiating music instruments, e.g. C7 played by a viola as opposed to C7 by a harpsicord. In speech, spectral changes produced by varying configurations of the vocal tract and other speech articulators result in different formant frequencies and transitions to distinguish different phonetic contrasts, most noticeable for vowels (Coleman, 2006; Gramley, n.d.) and phonemes with different places of articulation. For example, two vowels /u/ and /a/ may be produced at the same pitch with the exact intensity, yet one may still be distinguished from the other due to differences in sound quality (Koelsch, 2011).

Studies have found that musicians are more accurate than nonmusicians in processing spectral cues for both nonspeech (Seither-Preisler et al., 2007; Chartrand & Belin, 2006; Crummer, Walton, Wayman, Hantz & Frisina, 1994) and speech. They could better discriminate formant changes in speech sounds (Zuk et al., 2013) as well as vowel changes (Bidelman & Alain, 2015; Kempe, Bublitz & Brooks, 2015; Bidelman, Weiss, Moreno & Alain, 2014; Elmer, Hänggi, Meyer & Jäncke, 2013; Kühnis, Elmer, Meyer & Jäncke, 2013; Dworkis, 2012; Marie, Delogu, Lampis, Belardinelli & Besson, 2011; Gottfried & Xu, 2008).

1.3.3 Duration

For duration, positive transfer effects have also been documented. Duration, another acoustic dimension in music and speech, lends rhythm and meter to sound sequences. Short and long sound units are typically accompanied by weak and strong stress to reflect the relative prominence of sound in a periodic pattern. This prominence is physically correlated to timing, harmonic and/or intensity differences. Groupings of short-long, weak-strong sequences called metrical structures serve as organizing landmarks for melody in music and for syllables in speech (Katz, Chemla & Pallier, 2015). Speech sound units may be organized into periodic patterns with corresponding “beats” or linguistic prominence as a result of intensity and duration. These are denoted by the

presence of long vowels, geminates, consonant clusters, or vowel quality (Hayes, 1995, pp. 5-7; Fry, 1955). In syllables, relative prominence is perceived as stress where strong syllables are stressed and weak syllables unstressed. Groups of strong and weak syllables form metrical feet, and metrical feet in turn construct larger periodic patterns to lend speech cadence. Languages have been typologically organized by linguistic rhythm and categorized as stress-timed, syllable-timed, or mora-timed based on prosodic structure rules that determine where linguistic prominence is assigned (Nespor, Shukla & Mehler, 2011, 1147-1159).

At the phrase level, stress can place pragmatic value on certain words by emphasis of the semantic relevance and connotations associated with the word(s) in focus. At the lexical level, for a number of languages, it reflects phonemic contrast, such that the same word with differently stressed syllables would have different meanings. Hence, detecting duration differences at phrase and word levels is an important part of being competent in a language.

Musicians' extensive exposure to rhythmic regularities hones their sensitivity to timing variations in contrast to nonmusicians (Güçlü, Sevinc & Canbeyli, 2011; Rammsayer & Altenmüller, 2006; van Zuijen, Sussman, Winkler, Näätänen & Tervaniemi, 2005). This perceptual advantage has been associated with boosted speech segmentation skills (François, Chobert, Besson & Schön, 2013) and enhanced processing of linguistic metrical structures to detect duration differences at phrase and sentence levels (Magne, Jordan & Gordon, 2016; Zhao & Kuhl, 2013; Marie, Magne & Besson, 2011; Kolinsky, Cuvelier, Goetry, Peretz & Morais, 2009).

Duration changes occur more rapidly at the syllabic level to demarcate phonetic boundaries, allowing listeners to tell one contrast from another. In relation to a music-related transfer, two studies found musicians better at identifying differences in vowels with contrastive duration. Adult musicians with long-term music training demonstrated enhanced discrimination of nonnative Thai vowels of different durations (Cooper, Wang & Ashley, 2017) while 8-10-year-old children who were provided short-term music training of 12 months displayed significantly greater brain sensitivity to native vowel duration differences than nonmusician children (Chobert, François, Velay & Besson, 2014). However, to note, vowels with contrastive duration are not as prevalent in many languages as phonetic consonant contrasts which contain very small duration changes, measured in the onset and offset of articulatory movements, such as vocal fold vibration. These minute changes require finer perceptual abilities for successful discrimination and categorization of contrasts.

Relatively few studies evaluate the influence of music experience on detecting transient duration speech features given that transfers at this level are less intuitive. A few studies on segmental contrasts that involved subtle timing differences given that the unfamiliar contrasts could prove a perceptual challenge to listeners. Sadakata & Sekiyama (2011) were among a few to investigate the possibility of a far transfer. Their study showed correspondence between music experience and increased sensitivity to Japanese geminate stops with small duration differences. Three other studies examined the discrimination of voice onset time (VOT), reporting differences between musicians and nonmusicians. Before we discuss these studies, we elaborate on VOT.

VOT marks the interval where a stop consonant is released and when voicing begins, e.g. the vocal folds vibrate. Stop production begins with an articulatory closure, a release of the consonant usually followed by a vowel in a syllable. VOT values are determined in relation to the stop/consonant release as illustrated in Figure 1. Negative VOT is derived when voicing occurs before the release while positive VOT occurs after the release. Voicing initiated at the moment of release is denoted as zero VOT (Figure 1).

Figure 1. Schematic of release time and voice onset time

Release is often signaled by a short burst of air and in some cases accompanied by a longer period of air release resulting in a voiceless fricative-like noise known as aspiration before the onset of a vowel in a syllable (see Figure 2). Aspiration can be found in stops (Moosmüller & Ringen, 2004; Kohler, 1979), certain affricates (Li, Oh, Shao & Shuai, 2012) and fricatives (Zhu & Chen, 2016; Li et al., 2012; Jacques, 2011). VOT duration in aspiration varies according to phonology (Cho & Ladefoged, 1999; Volaitis & Miller, 1992; Lisker & Abramson, 1964) or place of articulation (Peterson & Lehiste, 1960; Fischer-Jørgenson, 1954).

Figure 2. Depiction of aspiration in a speech waveform. Adapted from Waveform (amplitude as a function of time) of the English word "above," by COMDJ, 2009, <https://commons.wikimedia.org/wiki/File:Waveform-above.png>. Reprinted courtesy of the copyright holder under Creative Commons License CC BY-SA 3.0

Within a language, VOT differences can signal phonemic contrasts and are regarded as an acoustic device to categorize stops. VOT is also indicated to be effective for distinguishing phonemic categories in cross-linguistic comparisons (Cho & Ladefoged, 1999; Keating, Linker & Huffman, 1983; Lisker & Abramson, 1964). Perception of VOT change is a factor considerably studied in language processing and acquisition research given that learners often find it difficult to perceive VOT differences denoting unfamiliar phonemic categories.

To our best knowledge, three studies examined the effect of music experience on VOT perception (Kühnis, Elmer, Meyer & Jäncke, 2013; Martínez-Montes et al., 2013; Zuk et al., 2013). Kühnis et al. (2013) presented large and small VOT deviants in native syllables and reported group differences among musicians and nonmusicians to large VOT but not small VOT deviants. Martínez-Montes et al. (2013) likewise examined brain response to large and small VOT deviants with the exception that nonnative syllables and harmonic sound tokens were used. Their study found that while musicians were observed to show faster brain response to large VOT deviants compared to nonmusicians, there were no observed group differences for small VOT deviants and all other duration deviants. In Zuk et al.'s study (2013), musicians showed better discrimination of native synthesized syllables on a VOT continuum. While the conclusions from these findings are less clear, they imply far-reaching effects of music experience on processing duration contrasts at a small scale in speech sounds.

1.3.4 *Summary*

Given what we have reviewed so far, there are numerous studies which illustrate the positive effect of music experience on a range of language processing tasks and denoting a strong correlation between music training and enhance linguistic processing. Yet keeping in mind that correlation does not denote causation, longitudinal studies have also been conducted to eliminate the possibility that previous differences between musicians and nonmusicians could be affected by other factors. In one such study, nonmusician children assigned to music training showed enhanced brain response to subtle pitch and duration changes in speech and nonspeech stimuli compared to children assigned to non-music class, e.g. art class (Chobert, François, Velay & Besson, 2014; Moreno et al., 2009). The studies reinforce findings of positive music-language transfers, while other studies demonstrate more direct transfers of music experience to language processing in relation to pitch, timbre, and duration cues.

We now turn our attention to a few theoretical frameworks posited in the literature to account for how cross-domain positive transfers are possible. These will be discussed in light of empirical findings.

1.4 **Theoretical Frameworks on How Music Influences Language**

To date, the process by which music experience motivates an advantage in many linguistic functions remains a source of conjecture. In the literature, a number of explanations have been proposed, comparing music-language processing in relation to cognitive mechanisms.

1.4.1 *Top-Down and Bottom-Up*

The top-down and bottom-up perspective provides a conceptual basis for music transfer effects, specifically how music training contributes to auditory sensitivity and plasticity. Back of it all are cognitive and sensory mechanisms that interact in auditory perception (Kauramäki, Jaaskelainen & Sams, 2007; Allen, Kraus & Bradlow, 2000). These complex mechanisms are in some way enhanced at different processing levels by auditory experience. Music, particularly music training, entails extensive domain-general auditory practice and attention to complex sound structures, all of which provide a unique advantage. This advantage is then translated into neural proficiency. Given this, music experience has been found to facilitate rapid plasticity in short-term auditory perceptual learning and discrimination (Seppänen, Hämäläinen, Pesonen & Tervaniemi, 2013). It also effects long-term plasticity reflected in structural changes in musicians' brain anatomy as a function of years of experience (Imfeld, Oechslin, Meyer, Loenneker & Jancke, 2009; Schmithorst & Wilke, 2002).

This is shown in greater brain volume across musician groups, including increased grey matter volume in the Heschl's gyrus (Bermudez et al., 2009; Schneider et al., 2002), in brain regions responsible for motor-auditory processing (Hyde et al., 2009; Gaser & Schlaug, 2003) and also in the cerebellum (Abdul-Kareem et al., 2011; Hutchinson et al., 2003). There have also been findings of an increase in white matter pathways (Mathias, Adrian, Thomas, Martin & Lutz, 2010; Schlaug, Jäncke, Huang, Staiger & Steinmetz, 1995a) of musicians' brains, suggesting better brain connectivity and more efficient processing. Other evidences suggest an overlap of brain network and neural resources when processing music and speech and the use of domain-general processes (Peretz, Vuvan, Lagrois & Armony, 2015; Perrachione, Fedorenko, Vinke, Gibson & Dilley, 2013; Kraus & Chandrasekaran, 2010; Nicholson et al., 2003; Patel, Peretz, Tramo & Labreque, 1998). Top-down and bottom-up accounts take these cognitive processing mechanisms into consideration to explain how they are improved by music experience to allow for greater sensitivity to acoustic features.

Top-down account suggests that higher-order cognitive mechanisms shape lower-order sensory functions to impact auditory perception. More specifically, taking music in reference, it is hypothesized that auditory plasticity is shaped by top-down influences, i.e. music learning/experience, on cognitive processing mechanisms which in turn allows for bi-lateral sensory advantage in perception across domains. Common cognitive mechanisms are proposed to be recruited to process music and speech. Extensive auditory experience afforded by long-term music training is thought to act as a tuning device for sound encoding by sharpening higher-level processes such as attention, working memory, and learning (Kraus & Chandrasekaran, 2010; Kraus, Skoe, Parbery-Clark & Ashley, 2009) which in turn modulate lower-level functions in the cochlear (Bidelman, Schug, Jennings & Bhagat, 2014; Perrot, Michey, Khalfa & Collet, 1999) and brainstem (Zhu, Xia & Shinn-Cunningham, 2011; Parbery-Clark, Skoe, Lam & Kraus, 2009a; Parbery-Clark, Skoe & Kraus, 2009b; Tzounopoulos & Kraus, 2009; Musacchia, Sams, Skoe & Kraus, 2007). However, top-down influence of music experience does not automatically develop auditory functions across board. Only certain important processes are enhanced while redundant features are inhibited by music experience to attenuate neural response for greater auditory acuity. In other words, a highly trained auditory system results in selective perceptual advantage across music and speech domains (Kraus & Slater, 2015, pp. 212-214; Strait & Kraus, 2011). Categorical perception may thus be influenced by top-down effects in a specific domain to process shared acoustic features in another domain. Music training enhances higher-level processing to develop categorization skill in music (Burns & Ward, 1978; Blechner, 1977; Locke & Kellar, 1973) which also develops finer

perceptual boundaries in speech as shown in differences between musician and nonmusician groups (Weidema, Roncaglia-Denissen & Honing, 2016; Elmer et al., 2014).

Bottom-up processing is likewise posited to contribute to positive transfers. Basic sensory cognitive mechanisms such as the cochlear, brainstem, midbrain nuclei and auditory cortex, are responsible for detecting subtle variations in stimulus features, e.g. intensity, reverberation, spectrotemporal cues in acoustic signals before passing to higher-level processes for further interpretation. Thus, lower-order mechanisms form a shared functional base shaped by enriched acoustic environments to act as systematic “machines” for processing salient auditory signals. Given that music experience involves extensive training and a greater degree of exercising of the auditory system than is imposed by speech, it is posited to strengthen lower-level mechanisms for sound encoding.

Past studies suggest a close interaction of top-down and bottom-up approaches in mechanisms engaged for music and language processing where cognitive functions at higher and lower levels are enhanced (Angenstein, Scheich & Brechmann, 2012; Tervaniemi et al., 2009). Musicians have been found to possess better working memory than nonmusicians (Suárez, Elangovan & Au, 2016). Music-trained adults (Franklin et al., 2008; Brandler, 2003; Chan, Ho & Cheung, 1998) and children (Ho, Cheung & Chan, 2003) have also demonstrated increased verbal memory. In other studies, musicians employed a more efficient memory updating process (George & Coch, 2011) and working memory strategy to store nonverbal auditory information (Schulze, Zysset, Mueller, Friederici & Koelsch, 2011). In relation to attention, musicians have demonstrated better inhibition of irrelevant auditory information (Kaganovich et al., 2013) and improved auditory attention (Zhu, Xia & Shinn-Cunningham, 2011; Strait, Kraus, Parbery-Clark & Ashley, 2010).

Music experience have also been shown to modulate lower-level subcortical functions. Behavioral and imaging studies have reported increased brain activation in musicians’ primary and secondary auditory cortices during mental auditory imagery (Bunzeck, Wuestenberg, Lutz, Heinze & Jancke, 2005; Kraemer, Macrae, Green & Kelley, 2005; Jancke & Shah, 2004; Yoo, Lee & Choi, 2001), a more effective performance in music-trained auditory brainstems, and improved neural encoding of auditory tokens of music and speech tokens (Parbery-Clark, Tierney, Strait & Kraus, 2012; Parbery-Clark, Strait & Kraus, 2011; Bidelman & Krishnan, 2010; Chandrasekaran, Krishnan & Gandour, 2009; Musacchia, Strait & Kraus, 2008; Dees, Russo, Wong, Kraus & Skoe, 2007; Musacchia, Sams, Skoe & Kraus, 2007). Musicians were also observed to possess more robust subcortical responses to acoustic and speech sounds, suggesting better perceptual representation defined by training (Strait, Parbery-Clark, Hittner & Kraus, 2012; Bidelman, Gandour & Krishnan,

2011a; Bidelman, Gandour & Krishnan, 2011; Bidelman, Krishnan & Gandour, 2011b; Bidelman & Krishnan, 2010; Lee, Skoe, Kraus & Ashley, 2009; Strait, Kraus, Parbery-Clark & Ashley, 2010). In addition, studies have revealed that non-auditory higher-level cognitive processes, such as learning, are improved by extensive auditory exposure, proposing that greater auditory experience hones a more sensitive ‘ear’ (Skoe, Krizman, Spitzer & Kraus, 2013; Engineer et al., 2004).

1.4.2 A Statistical Learning Model

An alternative perspective have proposed that music-related transfers are possible as a result of domain general statistical learning. The OPERA model (Patel, 2011; 2012; 2013) explains that music-related positive transfers are determined by shared neural networks and processing demands across music and language with focus on factors in music training that influence adaptive plasticity in speech processing. In contrast to the notion of shared cognitive processes in the top-down bottom-up framework, the model assumes an underlying domain-general sound learning mechanism for music and speech. Five conditions are specified in order induce neural plasticity for higher precision in encoding and perception of acoustic sounds: 1) an **overlap** in brain processing networks for music and speech; 2) a greater processing **precision** required for music than for speech due the greater demands music training places on the auditory processing networks; 3) positive **emotion** that accompanies music experience in order to engage processing mechanisms; 4) requisite music practice or **repetition**; and finally 4) focused **attention** on the detail of musical tokens.

Some of these conditions have indeed been shown to shape auditory processing. It was found that in an experimental group of infant and children participants, active music participation and consistent music practice resulted in larger cortical responses to tones compared to passive music listening and inconsistent music practice (Trainor et al., 2012) highlighting the significance of repetition and active attention as part of music experience. However, empirical findings have not verified if indeed all hypothesized conditions of Patel’s framework must be present in order for a speech processing advantage to result. To note, according to the hypothesis, should any one of OPERA conditions be absent, positive transfers from music training/experience would not be predicted (Asaridou & McQueen, 2013). This would also mean that musically-trained individuals who did not willingly participate, e.g. lacking positive emotion, and those who lack attention or practice in the course of music training would perform no differently from nonmusicians (Patel, 2011). Another point of consideration is the fact that the specified conditions are not unique to music experience and may be applied to certain cases of language learning, e.g. tone languages, thus suggesting that conditions may benefit neural networks on a broader scope to process relevant auditory information.

To account for this, a more specific version of statistical learning known as distributional learning have been proposed (Ong, Burnham, Stevens & Escudero, 2016). In this paradigm, knowledge acquisition is largely determined by successful detection of distributional structure of the incoming information. As example, Ong et al. (2016) discussed how distributional learning may influence perceptual attunement of voicing categories across different languages. Given a continuum of -120 ms to +20 ms, English infants exposed only to speech sounds modeled around a continuum with a single peak approximating 0 ms will likely develop a single voicing category. In contrast, Hindi infants with exposure to more nuanced speech sounds on the continuum will form two voicing categories modeled as two distributions with two peaks on the continuum. Past studies support this paradigm, demonstrating that distribution learning have been used to detect familiar information in unfamiliar contexts in the case of infant phonetic learning (Yoshida, Pons, Maye, & Werker, 2010) and learning unfamiliar vowel contrasts (Escudero, Benders & Wanrooij, 2011). In addition, distributional learning have also been found to contribute to acquiring unfamiliar information within the same domain (Hall, Owen Van Horne & Farmer, 2018; Escudero & Williams, 2014) and also cross-domains (Ong et al., 2016).

Taken together, top-down bottom-up and statistical learning models provide functional frameworks to facilitate our understanding cross-domain transfers.

1.5 Research Gap

While past studies provide good evidence for positive music-related transfers, it is clear that a majority only concern transfer effects to prosodic speech processing, viz., pitch, intonation and linguistic stress. They largely disregard transfer effects to segmental speech processing. Furthermore, there is a lack of direct comparison of specific acoustical features similar across music and segmental speech sounds. Of a small handful of studies on this topic, the set of stimulus category examined was often restricted primarily to vowels, limiting the generalizability of findings (Cooper, Wang & Ashley, 2017; Chobert, François, Velay & Besson, 2014; Kühnis, Elmer, Meyer & Jäncke, 2013; Reinke, He, Wang & Alain, 2003). Moreover, a majority examined perceptual sensitivity to native syllables which did not present a considerable challenge to participants (Chobert, François, Velay & Besson, 2014; Kühnis, Elmer, Meyer & Jäncke, 2013; Reinke, He, Wang & Alain, 2003). Whether music-related perceptual sensitivity may extend beyond the native language domain is less known.

To date, only two cross-linguistic studies have looked at possible interactions between music and nonnative segmental processing. They present mixed results. In a behavioral discrimination of duration changes in nonnative vowels and geminate stops, musicians were found to clearly

outperform nonmusicians, detecting differences as minute as 15 ms apart (Sadakata & Sekiyama, 2011). However, another study did not discover significant group differences across brain responses to VOT in nonnative syllables with comparable duration differences to those in geminate stops in the earlier-mentioned study (Martínez-Montes et al., 2013). The difference in findings may have arisen due to testing methods where distinctions could be better perceived at a behavioral level compared to a preattentive one, indicating possible disassociations between active and automatic processing. Moreover duration perception in the two studies may not be completely comparable owing to differences in processing mora-timed duration (geminate) and voicing duration. In response to the abovementioned points, our present investigation proposes to address some of the research gap and add to the insufficient number of studies in this field of study.

1.6 The Study

Taking the two earlier studies in consideration, the current study is conducted to contribute further information on possible music-related transfers to segmental speech beyond the native language domain. It examines if positive transfers to sensitivity to nonnative VOT contrasts with minute duration differences comparable to those in the geminate contrasts examined by Sadakata & Sekiyama (2011) given in a carefully controlled test design in view of the findings by Martínez-Montes et al. (2013). This study also seeks to analyze music experience with respect to separate components of aptitude, training, exposure, and to a lesser degree, instrument specialization, investigating possible influence of these factors on nonnative speech perception. The next section will present our research questions followed by a description of experimental design.

1.7 Research Questions

Considering past studies and the theories on music-transfer effects, it would be of interest to determine the impact of music experience on the ability to actively perceive nonnative phonemic contrasts with minute contrastive duration and timbre features. More specifically, we will examine perception of differences in voice onset time (VOT) and place of articulation (POA), particularly the dental-retroflex contrast, identified by past research as particularly challenging for nonnative speakers.

Across languages, VOT has been widely used to denote category boundaries for phonetic contrasts. Adult learners often find it difficult to perceive subtle VOT differences that denote unfamiliar phonemic categories in nonnative languages, making sensitivity to VOT change a topic of relevance in many language acquisition studies. Another difficult contrast is Hindi dental-retroflex /t-ɭ/ which ranks high on the difficult contrast list for nonnative listeners (Pruitt, Jenkins & Strange, 2006; Rivera-Gaxiola, Silva-Pereyra & Kuhl, 2005; Rivera-Gaxiola, Csibra, Johnson & Karmiloff-

Smith, 2000; Polka, 1991) and occurs in only 11% of the world's language (Golestani & Zatorre, 2004). It has been found that even after exposure and standard training, nonnative listeners, particularly English native speakers, did not show marked improvements, retaining the tendency to assimilate both sounds as allophones of the dental /t/ (Tees & Werker, 1984). Should music experience enhance perception of these contrasts, it would provide an important aspect to consider as a viable way to override perceptual difficulty in processing these challenging sounds.

It would also be of interest to examine different components of music experience and their effect on speech processing. This differs from previous studies where music experience typically denotes music training. In a majority of these studies, music-related transfers across domains are examined with respect to active music training, differentiating performance according to its presence or absence, instead of evaluating transfers from a more encompassing angle to include both active and passive music experience –something which our study proposes to examine.

Apart from music training, a study by Martínez-Montes et al. (2013) cited music exposure as a possible affective factor in perceptual discrimination. The authors reported a smaller-than-expected group difference between musicians, and their findings were attributed to nonmusicians' music exposure, as nonmusicians in the study were visual artists who listened to long hours of music while creating art pieces. This explanation is plausible given the bottom-up account where extensive exposure and experience is thought to hone one's auditory acuity. We thought to examine if music exposure would result in measurable perceptual differences. In a number of early studies, music aptitude was used to distinguish musician from nonmusician groups (Magne, Jordan & Gordon, 2016; Kliuchko et al., 2015; Strait, Hornickel & Kraus, 2011; Milovanov, Huotilainen, Valimaki, Esquef & Tervaniemi, 2008). For our study, we evaluated music aptitude in relation to perceptual accuracy across both musicians and nonmusicians. We also briefly considered instrument expertise, a variable explored in recent studies. These various components analyzed separately from music training afford a better look at specific factors which could motivate music-to-speech transfers.

We designed our study to examine these effects, exploring possible transfers in light of higher-level processes, e.g. auditory working memory, attention. Our research questions are as follows:

1. Will music experience enhance the perception and categorization of nonnative segmental VOT and dental-retroflex contrasts?
2. Apart from music training, will music aptitude, exposure, and instrument expertise result in performance differences?

To address these questions and shed light on the influence of music experience on nonnative segmental speech processing, two experiments were conducted. The first assessed whether music experience enhanced processing nonnative syllables with contrastive VOT, while the second investigated the effect of music experience on processing nonnative syllables with challenging place of articulation (POA) difference, specifically dental-retroflex contrasts.

Based on previous work, we hypothesized that music training will positively motivate music-related transfers in processing nonnative segmental tokens contrasting in VOT and POA. We also hypothesized that music aptitude and exposure will play in role to affect performance. We predicted that musicians would outperform nonmusicians in perceptual accuracy for discriminating and categorizing target contrasts; while participants with high music aptitude scores and greater daily music exposure than the average would also show differences from those with lower music aptitude and daily music exposure.

2 METHODS AND STATISTICAL ANALYSIS

2.1 Exploratory Study

The main study was conducted to examine musician and nonmusician processing of nonnative contrasts. Experiment 1 evaluated perception of nonnative voicing contrasts. We selected Hindi voicing contrasts to create test tokens, given that it has four-way contrasts which also pose a challenge for nonnative listeners. In order to ensure that target contrasts were unfamiliar and not meaningfully contrastive to participants, they were compared to those in participants' native languages. A small exploratory study was conducted to gauge voicing contrast distinction among native Singaporean English speakers. Before reporting this pilot study, we briefly discuss voicing categories in Hindi and participants' native languages, e.g. English, Malay and Mandarin Chinese.

2.1.1 *Voicing Categories*

As mentioned previously, voice onset time (VOT) marks the interval from release of a stop occlusion to the onset of glottal vibration that is often contrastively used to characterize stop consonants across many different languages. Voicing categories are determined by the onset of voicing with respect to the release burst in stops. Voicing before the release burst or “voicing lead” would result in prevoicing, denoted by negative values, e.g. [-VOT]. Voicing that occur after the release burst, “voicing lag,” are given positive values, e.g. [+VOT]. Lag is further subdivided into **short-lag**, 0 or low positive VOT, and **long-lag**, high positive values, e.g. over 35 ms, denoting aspiration (Lisker & Abramson, 1964; Keating, 1984).

2.1.1a Voicing Categories in Hindi

Hindi contains a four-way contrast for stops as seen in Table 1. By convention, voicing categories are defined by a single phonetic dimension, VOT, and acoustically realized in laryngeal contrast of two acoustic dimensions, voicing and aspiration. The four categories are prevoiced/voiced unaspirated [-VOT], short-lag/voiceless unaspirated [~ 0 VOT], long-lag/voiceless aspirated [+VOT], and prevoiced long-lag/voiced aspirated. Three of the stop categories are immediately distinguishable by VOT alone –the prevoiced, short and long lag. However there is some perceptual ambiguity for the prevoiced long lag or voiced aspirated stop, as its production involves

the simultaneous implement of voicing and aspiration at the release of the stop closure, where vocal cords are drawn together for voicing at the back while the front remains open to allow passage of large volumes of air to be indrawn, resulting in its characterization as a “breathy” sound (Hauser, 2016; Dutta, 2007).

Table 1

Hindi Stop Contrasts in Monosyllabic Tokens

	Bilabial	Dental	Retroflex	Velar	Palaatal
Voiceless unaspirated	/pa/ पा	/ta/ ता	/ʈa/ टा	/ka/ का	/tʃa/ चा
Voiceless aspirated	/pha/ फा	/tha/ था	/ʈha/ ठा	/kha/ खा	/tʃʰa/ छा
Voiced unaspirated	/ba/ बा	/da/ दा	/ɖa/ डा	/ga/ गा	/ja/ जा
Voiced aspirated	/bha/ भा	/dha/ धा	/ɖha/ ढा	/gha/ घा	/jʰa/ झा

2.1.1b Voicing Categories in English

In contrast, English possesses two voicing categories, the voiced and voiceless at initial syllable or word position. An example of these contrasts would be the words “bin” /bɪn/ and “pin” /pɪn/. There are two allophonic variations for voiced stops, either as lead/prevoiced [-VOT] or short lag/unaspirated [~0 VOT] stops (Keating, Linker & Huffman, 1983) where prevoicing is present in certain varieties such as southern American English (Hunnicutt & Morris, 2016; Jacewicz, Fox & Lyle, 2009) and British English (Lisker & Abramson, 1964). Thus, lead and short lag stops are not meaningfully contrastive (Jacewicz et al., 2009) in English. However lead (prevoiced) [-VOT] and short lag (unaspirated) [~ 0 VOT] phonemically contrasts with long lag (aspirated) stops [+VOT].

2.1.1c Voicing Categories in Mandarin Chinese and Malay

Voicing categories in Mandarin Chinese and Malay were taken into account to ensure that only unfamiliar nonnative voicing contrasts were used as test tokens. In Mandarin Chinese, stops are phonemically voiceless and stop contrasts categorized by aspiration. The phonological inventory contrasts voiceless short lag stops [~0 VOT] such as in 白 /pái/, meaning “white,” with voiceless long lag stops [+VOT] as in 拍 /pʰāi/, meaning “clap” (Chao, Peng, Yang & Chen, 2008; Duanmu, 2000). This voicing category is phonetically similar to the one in English.

In Malay, stops are categorized in relation to prevoiced [-VOT] and voiceless short lag contrasts [~0 VOT] given that there is no aspiration recorded in Malay phonology (Shahidi & Aman, 2011). An example of the [-VOT] and [~0 VOT] contrast would be the words /buak/ which means “effervescence” with a prevoiced stop, and /puak/ or “clan” with a voiceless unaspirated stop. Table 2 gives a summary of phonetic contrasts in relation to VOT for voicing categories across the four languages.

Table 2

Voicing Categories in Malay, English, Mandarin Chinese, and Hindi

	-VOT Pre-voiced/Lead	~0 VOT Short lag	+VOT Long lag	Voiced Aspirated Lead +Long lag
Malay	+	+		
English		+	+	
Mandarin Chinese		+	+	
Hindi	+	+	+	+

2.1.1d Voicing Categories in Singapore Language Varieties

In multicultural Singapore, a hodgepodge of linguistic codes and “dialects” coexist in a complex social, political, racial, and cultural setting. Language contact among speakers in this ethnic-linguistic melting pot is typical. Hence, different main languages, including English, Mandarin Chinese, and Malay have been infused with borrowings in lexicon, phonology, and even syntax across the various languages to develop a unique variety of Singaporean languages that differ from that those spoken in other places. These varieties are very much in use by a large majority of Singaporeans and in fact serve as a linguistic design to distinguish natives from speakers of other lands.

A local variety of English spoken by natives is Singapore English. This variety is comparable to standard English with speech features and phonology influenced by the daily contact with speakers of different ethnic groups. It was once observed that “it is no longer possible to tell a Chinese, a Malay and an Indian Singaporean apart from each other, just by listening to them speaking English” (Platt & Weber, 1980, p. 152). In relation to VOT, a few studies have found that Singapore English stops are produced with less aspiration than that of standard British or American English (Tay, 1993; Tee, 1986). Based on observation, a majority of Singapore English speakers appear to use short-lag stops [~0 VOT] to denote contrasts classified as voiced. To evaluate this observation, a pilot production study was conducted on a small sample of Singaporean native English speakers.

2.1.2 Participants

Twelve native Singaporean Chinese-English bilinguals (7 females) with an age range of 19 to 61 years (mean age = 32.4 years) participated in a pilot production study. Younger and older adults were included in this random sample as an exploratory gauge on VOT production in stops. All participants grew up in Singapore since birth and reported neurological or speech impairments which could possibly affect articulation. One participant below the age of 21 years participated with parental consent. Table 3 presents a list of languages known each participant. To note, languages known to the participants did not appear to influence their productions.

Table 3

Participant Language Background Information

	Participant	Languages known/spoken
1	10310	English, Mandarin Chinese
2	10311	English, Mandarin Chinese
3	10312	English, Mandarin Chinese
4	10314	English, Mandarin Chinese, Teochew
5	10317	English, Mandarin Chinese, Hokkien, Teochew
6	10319	English, Mandarin Chinese, Malay
7	10320	English, Mandarin Chinese, Cantonese
8	10321	English, Mandarin Chinese, Teochew
9	10313	English, Mandarin Chinese, Spanish
10	10315	English, Mandarin Chinese, Hokkien
11	10316	English, Mandarin Chinese, Hokkien, Teochew, Cantonese
12	10318	English, Mandarin Chinese

2.1.3 Methods

2.1.3a Materials

A list of monosyllable English minimal word pairs with CV(CC) structure was created containing phonemically voiced and unvoiced labial stops, /b/ and /p/ at word onset. The stops appeared in the environment of the vowels/diphthong namely /ɑ:/, /ɒ/, /ɪ/, /ʊ/, /eɪ/. Five minimally contrastive pairs were created resulting in a total of 10 tokens (see Figure 3).

1	bar	par
2	bull	pull
3	beige	page
4	big	pig
5	bond	pond

Figure 3. Word list of minimally contrastive stop tokens in pilot study

2.1.3b Task

Participants were recorded in a quiet room where they sat in front of a computer. They were asked to say each of the target words twice in the carrier phrase, “Say _____ again” at a regular speaking rate. Target words were pseudorandomized to form two list versions. Lists were counterbalanced across participants, such that six participants read from list version one while the other six read from list version two. Participants were recorded with the help of a high-quality microphone and the digital recording software Acoustica 6.0. Before the recording, participants were informed that the pilot study was conducted to investigate Singapore English. They were allowed to ask further questions about the purpose of the experiment after the recording.

2.1.3c Measurement

Acoustic measurements of recordings were made using Praat. All tokens were digitalized at 48 kHz and normalized at 70 dB. Target words were digitally extracted from the carrier phrases, and VOT values were manually measured by taking the interval between the onset of the release burst and beginning of the visible F1 on the spectrogram as shown in Figure 4.

Figure 4. Example of measurement method for VOT values

2.1.4 Results

It was found that overall the voiced labial stop /b/ had an average of 17.9 ms (SD = 6.41 ms) while the average VOT value for voiceless aspirated labial stop /p^h/ was 66.1 ms (SD = 67.6 ms). Productions that did not reflect the overall trend across participants were excluded from the analysis. These include productions of the aspirated stop /p^h/ by a speaker which contained marked aspiration and distinctly longer VOT values than others produced by participants. Recordings of four participants were analyzed separately due to consistent production of prevoicing [-VOT] for the voiced labial stop /b/ (M = -101.5 ms, SD = 23.0 ms), unlike eight other participants. It is acknowledged that production differences could in part be attributed to the age spread. Three out of four of the participants whose productions contained consistent prevoicing were above age 40 years. It is possible that the linguistic landscape during the older adults' formative years differs greatly from the formative linguistic landscape of the younger adults, e.g. 18-30 years, which could have led to production differences. Moreover, the phonetic environment where the stops occur also appeared to influence VOT values, such that certain productions of /b/ when followed by back vowels /ʊ/, /ʊ/ or the diphthong /eɪ/ resulted in a longer VOT values. Uncharacteristic VOT values produced in these cases were identified as outliers and excluded from the analysis. Refer to Table 4 for VOT values across productions.

Table 4

Production of Voiced and Voiceless Stops by Native Singapore English Speakers

Participant	VOT Values for Initial Stops (ms)*									
	bar	bull	big	bond	beige	par	pull	pig	pond	page
10310	14.2	15.3	12.0	8.70	17.3	149.	186.	229.	156.	126.
10311	13.9	15.7	12.6	14.5	79.7	83.7	43.0	29.7	67.1	39.2
10312	30.5	14.8	34.4	16.8	15.7	65.7	89.2	79.1	57.8	61.8
10314	14.2	34.7	16.3	17.1	17.0	68.2	57.7	31.4	42.1	41.1
10317	17.6	30.2	28.7	26.7	16.8	57.5	67.0	43.2	47.9	38.5
10319	28.7	24.5	20.6	37.0	15.4	34.0	75.3	48.6	32.3	43.2
10320	15.2	16.4	12.7	16.2	77.6	58.1	72.0	46.7	74.5	441.
10321	5.00	72.0	16.8	16.5	16.2	73.0	40.8	28.2	93.1	42.6
10313	-71.8	-92.3	-105.	-138.	-101.	53.8	61.2	34.3	65.4	49.0
10315	-102.	-113.	-67.3	-90.2	-108.	68.9	75.5	42.6	64.8	45.9
10316	-131.	-122.	-112.	-86.6	-116.	117.	88.2	50.6	90.4	78.9
10318	-128.	-101.	-46.8	-118.	-79.0	103.	78.3	46.1	111.	32.2

*Values in red were excluded from the analysis as they did not reflect the trend shown by the majority. Shaded areas denote productions that were analyzed separately as they contained prevoiced stops

Overall, the data suggested that while allophonic variations exist for voiced stops in Singapore English, there is a trend for voiced stops to possess VOT values classified as short-lag [~ 0 VOT]. Across eight participants in the analysis, voiced /b/ and voiceless /p/ were distinguished by producing either [~ 0 VOT] or [+VOT] values.

2.1.5 Summary

The results from the pilot production study may be examined parallel to findings by Ng (2005) on VOT in stops across five Singaporean native languages –Mandarin Chinese, Malay, Tamil, Hokkien, and Cantonese. The study suggested that there is no clear boundary between [-VOT/0 VOT] and [+VOT] stops across the five languages. It also found that although voiced and voiceless stops are phonemically contrastive in Malay, VOT difference was not acoustically reflected in the production of Singaporean Malay speakers, with the exception of velar stops (Ng, 2005, p. 101). Moreover, among Singaporean Malay, Chinese, and Tamil speakers, VOT values for short-lag stops [~ 0 VOT] were similar as was the case in values for long-lag stops [+VOT]. Taken together, these findings curiously imply two main voicing categories [~ 0 VOT] and [+VOT] used to denote voiced and voiceless stop contrasts across many of main languages spoken here. The findings, however, are by no means conclusive given variations in VOT contrasts and production in other local linguistic varieties across different ethnic groups, e.g. Peranakan English (Lim, 2010). Table 5 presents a visual summary of voicing contrasts across the Singapore language varieties discussed by Ng (2005). From the conclusions of this earlier study and our preliminary findings, we then identified unfamiliar voicing contrasts for Experiment 1 of the main study.

Table 5

Summary of Voicing Contrasts in Singapore Language Varieties Discussed by Ng, 2005

	-VOT Pre-voiced/Lead	~ 0 VOT Short lag	+VOT Long lag	Voiced Aspirated Lead +Long lag
Singapore Malay	(+)*	+		
Singapore English		+	+	
Singapore Mandarin		+	+	
Singapore Hokkien		+	+	
Singapore Cantonese		+	+	

*Note: [-VOT] is only phonemically contrastive to [~ 0 VOT] in Malay velar stops

2.2 Experiment 1

2.2.1 *Participants*

Eighty-eight young adults between 18-35 years were assigned into groups based on music background, 44 musicians (mean age = 20.8 years, 34 females, 2 left-handed) and 44 nonmusicians (mean age = 22.2 years, 24 females, no left-handed). Participants were recruited from three local universities, Nanyang Technological University, National University of Singapore, and Singapore Management University, for a two-hour experiment. Musicians were later evaluated according to primary instrument expertise –keyboard musicians (n=29), string musicians (n=9), and wind musicians (n=6). All participants were native English-Chinese and English-Malay bilinguals with no known hearing or neurological disorders and also no exposure to Hindi, the target language in the study. Participants completed a detailed questionnaire form on music training background, age of music acquisition, number of instruments played, years of music training/music band practice, current number of practice hours a week, educational and linguistic background. Refer to Appendix A for participants' reported language backgrounds. Participants were included in the study only when it was determined that test contrasts were not meaningfully contrastive in native and known languages. A number of participants reported some knowledge of Cantonese (N=9) and Hokkien (N=17). Much like Mandarin Chinese, both languages have been found to discriminate stops by the presence or absence of aspiration, in particular for Cantonese, which was found to present large VOT values due to aspiration (Hong, 2012; Ng, 2005; Tsui & Ciocca, 2000). A number of participants reported knowledge of other languages learned at some point of their lives that cannot be considered native, e.g. Spanish, German, Korean. Importantly, the stop contrasts in these languages were not thought to influence perception, and test contrasts remained unfamiliar to participants. All participants consented to participate in the experiment and received monetary compensation for their time according to guidelines provided by the NTU Institutional Review Board.

Musicians were then identified according to age of music acquisition and years of music training based on a review of previous music studies which show behavioral and neural plasticity transfer effects of music to speech processing (Cooper & Wang, 2012; Skoe & Kraus, 2012; Bidelman et al., 2011; Bidelman & Krishnan, 2010; Chandrasekaran et al., 2009; Parbery-Clark et al., 2009; Zendel & Alain, 2009). Musicians were further screened based on preliminary data in a pilot study which found musicians who practice for less than an hour weekly and nonmusicians who reported self-learning an instrument after 12 years of age showed marked performance differences from

other participants in their respective groups. Thus, the inclusion criteria for musicians in this study was set as those with a music acquisition onset no later than 12 years of age, having at least 4 years of formal music training or intensive band practice and practicing an instrument for at least two hours per week at the time of testing. Three musicians were of an advanced level of teaching music but none of the other musicians reported any other music endeavors, e.g. composing, lyric writing. Nonmusicians were to have no more than two years of music experience and not playing an instrument at the time of testing. This inclusion criteria was consistent with many earlier studies (Cooper, Wang & Ashley, 2017; Bidelman, Schug, Jennings & Bhagat, 2014; Zuk et al., 2013; Boh, Herholz, Lappe & Pantev, 2011; Bidelman & Krishnan, 2010; Chandrasekaran et al., 2009; Herholz, Lappe & Pantev, 2009; Parbery-Clark et al., 2009).

A majority of the nonmusician participants reported attending basic music introductory classes as a requisite part of the Singapore primary school program. This however was not considered to constitute formal music training. Nonmusicians did not have any other music experience beyond introductory music classes. Table 6 presents a summary of music experience across groups.

Table 6

Summary of Group Statistics

Demographics	Musicians			Nonmusicians		
	M (SD)	Min	Max	M (SD)	Min	Max
Onset of formal music training (age)	6.5 (1.9)	4	11	N.A*	N.A	N.A
Music training (age)	12.9 (5.0)	4	>10	N.A	N.A	N.A
Current practice (hours/week)	3.7 (2.6)	2	>10	N.A	N.A	N.A

*Nonmusicians only attended basic music introduction classes in primary school which is not considered to constitute formal music training.

2.2.1a Participant Exclusion

For all tests, a number of participants were who were initially included in the experiments as musicians but later reported formally acquiring music after 12 years of age or practicing an instrument less than an hour weekly were excluded. Musicians who scored below the normal range of the group for the Musical Ear Test were also identified as outliers and removed from the data analysis.

In addition, a number of nonmusicians who later reported self-learning an instrument after the age of 12 years, despite not playing the instrument at the time of testing were excluded from the participant group due to unpredictable results in the preliminary data where their performance fell neither in the range of typical nonmusician nor musician groups.

2.2.2 *Methods*

2.2.2a Materials

Five unfamiliar voicing contrasts in Hindi were selected, [-VOT vs. 0 VOT], [-VOT vs. +VOT], [+VOT vs. voiced aspirated], [-VOT vs. voiced aspirated], and [0 VOT vs. voiced aspirated]. Hindi stop contrasts in each of these categories were chosen as initial stops for test tokens. The stimuli consisted of a naturally-spoken token set containing 48 CV monosyllables with dental, retroflex, velar and palatal stops where the vowel was one of /a/, /e/, or /o/. Refer to Appendix B for list. All tokens were judged to be phonotactically-acceptable by two native Hindi speakers (1 male, 1 female). Tokens were then recorded by four other native Hindi speakers (2 males, 2 females) to capture within-category variability. Native speakers produced two instances of each token embedded in the Hindi carrier sentence एक _____ एक बोली, which is an equivalent of “Say _____ again” and were recorded in a sound-attenuated room with a Shure SM81-LC microphone (20 Hz–20 kHz frequency response) using the software Acoustica 6.0 at a 44.1 kHz sampling rate (32-bit resolution, mono). Production accuracy of each token was verified by a comparison across native speakers. Based on voice quality and clarity of enunciation, the recording of one male native Hindi speaker was selected as a prototypical version of the test tokens and used to create tests.

Test tokens were digitally extracted from the sentence frame, digitalized at 48 kHz and normalized at 70 dB using NCH software Wavepad Sound Editor. Minimal editing was performed to ensure that approximately equivalent syllable lengths across tokens, and mean duration of all CV tokens was 698 ms. Tokens were then used to form pairs with minimally contrastive voice onset time (VOT) and classified under one of the five voicing contrasts identified earlier: [0 VOT vs. -VOT], [-VOT vs. voiced aspirated], [0 VOT vs. voiced aspirated]. For consistency, tokens in a pair were separated by a 500 ms interval, while an 80 ms silence break was inserted before and after each syllable/or syllable pair.

After a preliminary study on 24 participants (12 musicians, 12 nonmusicians), the contrasts [0 VOT vs. -VOT], [-VOT vs. voiced aspirated], and [0 VOT vs. voiced aspirated] were found to be more

challenging, for all pilot participants. The other two contrasts, [-VOT vs. +VOT] and [+VOT vs. voiced aspirated] showed comparable group scores at ceiling performance and were deemed to be more perceptually salient than the other contrasts. These were excluded from the main experiment. Examples of syllable pairs for test contrasts are shown in Table 7.

Table 7

Voicing Contrast Examples

-VOT, 0 VOT	-VOT, Voiced aspirated	0 VOT, Voiced aspirated
de-te	da-dha	ka-gha
ɖa-ta	je-jhe	ce-jhe
ge-ke	ɖo-ɖho	to-ɖho
jo-co	ga-gha	ta-ɖha

Apparatus

A DELL desktop computer with IntelCore duo processor i5 (12 GB RAM) and 17.7-inch screen was used to present visual instructions and auditory tokens. The software ePrime 2.0 was used to run the discrimination tests and to collect responses.

2.2.2b Tasks

In a single two-hour test session, participants completed 1) pre-tests evaluating basic hearing ability, nonverbal intelligence, auditory working memory, auditory attention, and music sophistication and music aptitude; 2) and two perceptual experiments, an AX and ordered discrimination test. To prevent testing fatigue, short breaks were provided during each test along with a requisite five-minute rest interval between tests. Participants were seated comfortably in a sound-attenuated room where they listened to test tokens via Sennheiser HD 280 Pro headsets and responded by selecting appropriate keys on the computer keyboard quickly and accurately.

Audiometric Test

An air-conducted audiometric test ensure normal hearing levels across participants. A series of pure sine ranging from 500 Hz -4000 Hz at a threshold of 25 dB were presented twice each in the left and right ear. Only participants who detected all test tones in both ears were included in the main experiment.

The Musical Ear Test

To evaluate musical aptitude across participants, the Musical Ear Test (MET) by Wallentin et al. (2010) was used. The test is based on the assumption that music aptitude includes auditory memory and the ability to detect melodic (pitch and contour) and rhythmic variations in short piano sequence pairs. Participants listened to sequence pairs and judged if the pairs were identical. According to expectations, musicians scored significantly higher than nonmusicians ($M = 85.11$, $SD = 6.44$), $t(86) = 8.84$, $p < 0.001$, outperforming in both the melody, $t(86) = 10.61$, $p < 0.001$, and rhythm subtests, $t(86) = 3.77$, $p < 0.001$.

The Goldsmiths Musical Sophistication Test

Participants also completed the Goldsmiths Musical Sophistication Test version 1.0 primarily developed to evaluate music experience in nonmusician populations (Müllensiefen et al., 2014). Music sophistication is measured as an index across five subfactors, namely active music engagement, music perceptual abilities, music training, singing abilities, and emotional response to music. The test, a self-report questionnaire, consists of 38 items rated on a Likert scale and a few questions on participant background (Appendix C). The index serves as supplementary information to compare with musicians' performance on the MET, while providing a measure of music aptitude for nonmusicians. The total general sophistication score was computed from scores across each of the five subfactors using a provided scoring template. For our study, the test was also used to classify participants based on music exposure, e.g. the amount of daily active music listening.

A significant group difference was observed where musicians demonstrated greater music sophistication than nonmusicians across all subfactors –active engagement, $t(84) = 8.59$, $p < 0.001$; perceptual abilities, $t(84) = 8.07$, $p < 0.001$; music training, $t(84) = 19.11$, $p < 0.001$; emotions, $t(84) = 5.38$, $p < 0.001$; singing abilities, $t(84) = 6.73$, $p < 0.001$ – and for general music sophistication, $t(84) = 12.11$, $p < 0.001$.

Edinburgh Handedness Inventory

Consistency of handedness was recorded using an adapted form of the Edinburgh Handedness Inventory (Oldfield, 1971). Participants were asked to indicate their hand preference (left or right) in carrying out 12 manual actions, e.g. using a spoon. Participants were predominantly right-handed, with the exception of two individuals.

Other Measures

Auditory attention and auditory working memory were measured using subtests in the Woodcock-Johnson III Tests of Cognitive Abilities. Total raw scores for each subtest were analyzed for each participant. No significant group differences were observed for auditory working memory, $t(86) = 1.37$, $p = 0.174$, while a marginal significance was observed for auditory attention, $t(86) = 1.92$, $p = 0.058$. This marginal significance will be discussed later in relation to musicians' performance on Experiment 1. The Test of Nonverbal Intelligence, fourth edition (TONI-4) was also administered to control for between-group differences in cognitive capability. No significant group differences were found between musicians and nonmusicians, $t(86) = 1.44$, $p = 0.155$.

AX Discrimination Test

A speeded AX discrimination also known as a two-alternative forced choice (2AFC) test, was used to establish participants' sensitivity to nonnative voice onset time (VOI) contrasts in Hindi phonemes. Test pairs were randomized into six test blocks (three main blocks for each voicing contrasts, repeated twice) resulting in a total of 144 trials. Each block consisted of 24 trials (12 same, 12 different pairs) which lasted approximately 5 minutes. Participants were informed that they would hear two sounds in a pair and asked to judge if pairs were identical or not, by keying "s" if they thought the tokens were same, and "d" if they thought tokens were different. Participants were given 3000 ms to respond and could only begin keying responses when a prompt appeared. Participants were first presented a practice block containing 5 trials with native voicing contrasts. Feedback to responses was provided during practice. Test blocks were then presented in counterbalanced order across participants with no feedback given. Participants' scores were calculated as d-prime scores (MacMillan & Creelman, 2005) to evaluate the proportions of hit rates, e.g. correct discriminations, and false alarms, e.g. incorrect discriminations of identical tokens, given the equation $d' = Z\text{score}(\text{Hits}) - Z\text{score}(\text{False Alarms})$.

Ordered Discrimination Test

The second test primarily tested performance on ordered discrimination, or sequence recall, but also involved categorization tasks. The test was first introduced by Dupoux et al. (2001) and deemed to involve higher-level cognitive processing, particularly attention and auditory working memory. It was included as a comparison to the simple AX discrimination to ascertain if group differences would also be found. Monosyllabic tokens from the test set were used to create test pairs for each of the three voicing contrasts. A total of three test blocks were created, each containing six test pairs and sequences.

In the test, participants first underwent a categorization task. They learned to associate each token in the pair to the number keys [1] or [2]. For example, given a test pair /d̪a-t̪a/, the syllable /d̪a/ would correspond to the number key [1], and /t̪a/ corresponded to the key [2]. Participants could listen to each token as many times as they preferred by keying the number associated with the token. When they were ready to move on, participants were presented a short quiz to test if they had learned to categorize tokens. Individual tokens in the pair were presented in random order, and participants were asked to key the associated number, [1] or [2] quickly and accurately. Next, participants underwent an ordered discrimination test where tokens in a test pair were pseudo-randomly presented in sequence, e.g. /d̪a-t̪a-t̪a-d̪a-d̪a/. Participants replicated the sequence by keying numbers corresponding to each token in order of the sequence, e.g. 12211. In Dupoux's design (2001), performance differences were found across test sequence lengths of two, four, and six. For this test, sequences of five were selected to ensure an appropriate level of test difficulty in consideration of the use of higher-level processing. To diminish the likelihood of participants using recoding strategies (that is, mentally translating words into corresponding number during sequence presentation), the silent period between tokens in each sequence was kept short, i.e. 80 ms. Participants were only able to key responses when they saw the prompt "Key answer now" which appeared immediately after the presentation of each token or sequence in the quizzes.

A practice block of 10 categorization trials (2 native contrast pairs x 5) and 2 sequence trials was presented followed by three test blocks. Each test block contained six contrast pairs, e.g. 30 categorization trials (6 test pairs x 5 trials) and 6 sequence trials (1 trial per test pair). A total of three test blocks (90 categorization trials, 18 ordered discrimination trials) were presented in counterbalanced order across groups. Feedback was provided for the practice but not for test blocks.

Speech and Nonspeech Conditions

Further, the test was assigned to two exploratory conditions where participants were primed to activate either speech or nonspeech mode during the experiment. A study by Takayama (2003) found significant priming effects between participants who were instructed to process sine wave analogs as speech sounds and those who were instructed to process the exact same stimuli as computer sounds. We included speech and nonspeech conditions in our study to establish if a similar priming effect could be found when naturalistic speech syllables were presented, and if there were any group differences. Half of the participants in musician and nonmusician groups were randomly assigned to the speech condition where they were told that they would listen to words of a new language, while the other half of the participants in the groups were assigned the nonspeech

condition and told they were listening to alien sounds. Apart from the initial priming instructions, the test was identical across both conditions.

2.2.3 Results

2.2.3a AX Discrimination Test

Music Training

A two-way ANOVA (group x contrast) was performed to analyze the data for the AX discrimination test. Results revealed a significant main effect of group, $F(1, 264) = 76.77, p < 0.001$ and contrast, $F(2, 264) = 37.60, p < 0.001$. The interaction effect of group and contrast was not significant, $F(2, 264) = 2.45, p = 0.088$. Figure 5 illustrates group differences in d-prime scores where musicians were better able to detect nonnative VOT contrasts than nonmusicians.

Figure 5. Mean discrimination scores for VOT test contrasts across groups. Error bars represent standard error of the mean (± 1 SE).

The trend was also significantly reflected in performance scores in each of the three VOT contrast conditions as shown in Figure 6.

Figure 6. Musician and nonmusician average discrimination scores across voicing contrast sets. Correct responses across groups for **(A)** 0 VOT vs. -VOT syllabic contrasts, **(B)** 0 VOT vs. voiced aspirated syllabic contrasts, and **(C)** -VOT vs. voiced aspirated syllabic contrasts. Error bars represent standard error of the mean (± 1 SE).

Correlation analysis revealed that total years of music training ($r = .486$, $p < 0.001$), weekly practice hours ($r = .273$, $p = 0.028$), and music band experience ($r = .486$, $p < 0.001$) positively related to d' -prime scores for discrimination of voicing contrasts across groups.

Music Aptitude

A significant positive correlation was found between performance scores and general music sophistication ($r = .385$, $p < 0.001$) including the five self-reported subfactors: active music engagement ($r = .228$, $p = 0.035$), perceptual abilities ($r = .333$, $p = 0.002$), music training ($r = .482$, $p < 0.001$), emotions towards music ($r = .235$, $p = 0.029$), and singing abilities ($r = .335$, $p = 0.002$). Scores also positively correlated with performance scores on the Musical Ear Test ($r = .463$, $p < 0.001$) across melody ($r = .496$, $p < 0.001$) and rhythm ($r = .265$, $p = 0.012$) subtests.

Music Exposure

A two-way ANOVA (group x daily music listening) explored the effect of musicianship and music exposure on discrimination accuracy, where participants were categorized as either listening to less than an hour, or at least an hour of music daily. A main effect of group was found where musicians performed better than nonmusicians regardless of music listening time, $F(1, 86) = 25.57$, $p < 0.001$. The number of daily music listening hours across participants did not significantly correlate with discrimination scores ($r = .177$, $p = 0.103$).

Instrument Specialization

The impact of music competence in different instruments on performance was determined by a one-way ANOVA which found no statistical performance difference between keyboard, string and wind musicians performance scores on the task $F(2, 41) = 1.590$, $p = .216$.

2.2.3b Ordered Discrimination Test

Music Training

For the ordered discrimination test, raw scores were tabulated across participants. This differs from the original design where error percentages were calculated by a difference score; that is, the error percentage of familiar contrasts minus error percentage of unfamiliar contrasts (Dupoux et al., 2001). In our adapted design, only unfamiliar contrasts were tested and the total number of correct responses was used to as a measure of performance. Figure 7 presents the correct response rate for musicians and nonmusician in categorization of nonnative VOT contrasts.

Figure 7. Average categorization scores across musician and nonmusician groups. Error bars represent standard error of the mean (± 1 SE).

A significant main effect of group was observed demonstrating the influence of musicianship on nonnative VOT categorization accuracy, $F(1, 264) = 43.57$, $p < 0.00$. A main effect of contrast was also indicated, $F(2, 264) = 18.835$, $p < 0.001$ along with a significant interaction of group and contrast on categorization scores, $F(2, 264) = 4.15$, $p = 0.017$. Figure 8 shows participant categorization scores across groups for different contrasts. Successful categorization of VOT contrasts positively corresponded with years of continuous music training ($r = .416$, $p = <0.001$), number of years in a music band ($r = .416$, $p < 0.001$), and the number of practice hours a week ($r = .389$, $p < 0.001$) across musicians and nonmusicians.

Figure 8. Mean categorization scores of musicians and nonmusicians. Musicians show greater categorization accuracy than nonmusicians across **(A)** 0 VOT vs. -VOT syllabic contrasts, **(B)** 0 VOT vs. voiced aspirated syllabic contrasts, and **(C)** -VOT and voiced aspirated syllabic contrasts. Error bars represent standard error of the mean (+/- 1 SE).

For ordered discrimination, main effects of group, $F(1, 264) = 68.24$, $p < 0.001$, and contrast, $F(2, 264) = 6.17$, $p = 0.002$, were likewise found, with no significant interaction between group and contrast, $F(2, 264) = 1.70$, $p = 0.186$. Total group scores and group differences across contrasts are shown in Figure 9 and 10 respectively. Ordered discrimination scores also significantly correlated with music training years ($r = .526$, $p < 0.001$), band practice years ($r = .460$, $p < 0.001$), and weekly practice hours ($r = .280$, $p = 0.024$) across groups.

Figure 9. Average correct responses in ordered discrimination of all VOT contrasts across musician and nonmusician groups. Error bars represent standard error of the mean (± 1 SE).

Figure 10. Ordered discrimination scores across contrast sets. Musicians outperformed nonmusicians in accurately reproducing sequences of syllabic contrasts containing the following differences: **(A)** 0 VOT vs. -VOT, **(B)** 0 VOT vs. voiced aspiration, and **(C)** -VOT vs. voiced aspiration. Error bars represent standard error of the mean (± 1 SE).

Music Aptitude

Categorization performance was found to correlate with general music sophistication ($r = .493$, $p < 0.001$) across the five subfactors: active music engagement ($r = .284$, $p = 0.008$), perceptual abilities ($r = .493$, $p < 0.001$), music training ($r = .529$, $p < 0.001$), emotions towards music ($r = .220$, $p = 0.042$), and singing abilities ($r = .438$, $p < 0.001$). Additionally, categorization of VOT contrasts correlated positively with performance on the MET ($r = .524$, $p < 0.001$) for melody ($r = .540$, $p < 0.001$) and rhythm ($r = .333$, $p = 0.002$) subtests.

Ordered discrimination scores indicated that music aptitude ($r = .632$, $p < 0.001$), across melody subtest ($r = .601$, $p < 0.001$) and rhythm subtest ($r = .476$, $p < 0.001$); and sophistication ($r = .535$, $p < 0.001$), e.g. engagement ($r = .359$, $p = 0.001$), perceptual abilities ($r = .493$, $p < 0.001$), music training ($r = .578$, $p < 0.001$), emotions towards music ($r = .275$, $p = 0.011$), and singing abilities ($r = .431$, $p < 0.001$), were associated with better performance.

Music Exposure

Music exposure was not found to have a significant influence on both categorization and ordered discrimination in a two-way ANOVA with music listening and group as fixed factors. Hours spent daily listening to music did not correlate with performance on categorization ($r = .129$, $p = 0.237$) and ordered discrimination ($r = .061$, $p = 0.575$). A main effect of group was found for categorization, $F(1, 86) = 29.48$, $p < 0.001$ and ordered discrimination, $F(1, 86) = 33.00$, $p < 0.001$, instead.

Instrument Specialization

A one-way ANOVA found no statistical difference across keyboard, string and wind musician scores for the categorization task, $F(2, 43) = 1.015$. However, musicians performed differently for ordered discrimination, $F(2, 43) = .587$, $p = 0.006$. Keyboard musicians scored significantly higher than wind musicians ($p = 0.007$) according to post hoc comparisons using the Bonferroni test (see Figure 11). There were no significant performance differences for ordered discrimination scores between keyboard and string musicians ($p = .289$); and between string and wind musician ordered discrimination scores ($p = .394$).

Figure 11. Group differences across keyboard, string, and wind musicians in ordered discrimination of VOT contrasts. Keyboard musicians had a significantly higher score than wind musicians. Error bars represent standard error of the mean (± 1 SE).

Auditory Attention

Interestingly, auditory attention was found to correlate with categorization ($r = .242$, $p = 0.023$), even though there were no significant differences found between musicians' and nonmusicians' auditory attention scores in the initial screening tests. This finding suggests that attention is relied upon as a resource in the categorization task.

Speech and Nonspeech Conditions

Recall that speech and nonspeech conditions were included for this test as an exploratory factor. Participants across groups were randomly assigned to either condition in counterbalanced order. Those assigned to the nonspeech condition were told that they were listening to alien sounds while those in the speech condition were told they were listening to words of a new language. Besides this initial information, the rest of the test was identical across conditions.

A univariate ANOVA showed a main effect of group for categorization where regardless of test condition, musicians consistently categorized voicing contrasts better than nonmusicians, $F(1, 88) = 33.500$, $p < 0.001$. On the other hand, a significant main effect of group, $F(1, 88) = 39.974$, $p < 0.001$, and significant interaction between group and priming condition were found for the ordered

discrimination task, $F(1, 88) = 4.607$, $p = 0.035$. Musicians demonstrated higher accuracy when they perceived contrasts as alien sounds while nonmusicians performed better when they perceived contrasts as words of a new language depicted in Figure 12.

Figure 12. Mean ordered discrimination score as a function of music training and perception priming condition. Error bars represent standard error of the mean (± 1 SE).

2.3 Experiment 2

While Experiment 1 assessed musician and nonmusician perceptual discrimination and categorization of nonnative VOT contrasts in Hindi, Experiment 2 measured perceptual performance in processing nonnative place of articulation (POA) contrasts, namely the dental-retroflex contrast. The idea is based on findings reported in an old study by Tees & Werker (1984).

In the study, American English speakers were able to discern nonnative Hindi voicing contrasts within less than a year of training yet could not do so with dental-retroflex contrasts. This suggests that for nonnative speakers, perceiving subtle unfamiliar POA contrasts may be a greater challenge than perceiving subtle unfamiliar voicing contrasts. Experiment 2 was included to observe the potential effect of music experience on processing these difficult dental-retroflex contrasts.

2.3.1 Participants

The participants who took part in Experiment 1 also took part in an AX discrimination of dental-retroflex contrasts.

A follow-up test was conducted to investigate categorization of dental-retroflex contrasts. This test had a slightly different participant group, comprising 33 young adults –13 musicians (mean age = 20.0 years, 8 females) and 20 nonmusicians (mean age = 21.6 years, 14 females). Participants were all predominantly right-handed undergraduate students at the Nanyang Technological University. They were native English-Chinese and English-Malay bilinguals with no known hearing and neurological disorders and no previous exposure to Hindi.

2.3.2 *Methods*

2.3.2a Materials

AX Discrimination Test Stimuli

Tokens from the stimuli set in Experiment 1 were used to form dental-retroflex contrastive pairs. Test pairs were randomized into two test blocks, each block consisting of 24 trials (12 same, 12 different pairs) resulting in a total of 48 trials. All test blocks were counterbalanced across participants. A practice block of five trials preceded the test blocks, and feedback was provided for the practice only.

Categorization Test Stimuli

Tokens from the main stimuli set were used to create a set of 12 dental-retroflex contrast pairs, randomized into two categorization test blocks which were counterbalanced across participant groups. Each block contained six dental-retroflex contrasts resulting in 60 tokens, to make a total of 120 trials across two blocks. A practice block with two native contrasts for 20 tokens. Here, as in other tests, feedback was only provided in the practice block.

2.3.2b Tasks

AX Discrimination Test

Participants underwent a speeded AX discrimination of dental-retroflex contrasts. The test design was the same as the AX test in Experiment 1. Participants were presented test pairs and asked to determine whether two tokens in a pair were identical. Participants keyed “s” if they thought tokens were the same and “d” if they thought tokens differed. D-prime scores were then calculated for data analysis.

Categorization Test

A categorization test was also conducted. This test was similar to the categorization task in Experiment 1. As dental-retroflex pairs are acoustically challenging to discriminate, figures were included to highlight a difference between two sounds in a pair. Each token in a pair was associated with a numbered geometric figure. Participants keyed [1] or [2] as many times as they wished to listen to the sound token corresponding to each figure (see Figure 13). After the training phase, participants were presented a categorization quiz, where they were presented sound tokens and asked to key the number [1] or [2] to indicate the picture matched to each sound.

Figure 13. Schematic of categorization task.

Note: IPA notations were not included in the actual test

2.3.3 Results

2.3.3a AX Discrimination

Music Training

Mean group d-prime scores for musicians ($M = 0.966$, $SD = 0.374$) and nonmusicians ($M = 0.657$, $SD = 0.472$) were significantly different, $t(86) = 3.401$, $p < 0.001$, indicating superior performance of musicians in discriminating dental-retroflex contrasts, depicted in Figure 14. Furthermore, there was a significant positive correlation between scores and the number of continuous music training years ($r = .287$, $p = 0.012$) and years of band practice ($r = .254$, $p = 0.057$) but not for practice hours ($r = .139$, $p = .269$).

Figure 14. Mean total score in discriminating dental-retroflex contrasts across musicians and nonmusicians. Error bars represent standard error of the mean (+/-1 SE).

Music Aptitude

Perceptual discrimination of dental-retroflex contrasts correlated with the extent of music aptitude in relation to general music sophistication ($r = .290$, $p = 0.007$) across the following subfactors: perceptual abilities ($r = .271$, $p = 0.012$), music training ($r = .283$, $p = 0.008$), emotions ($r = .240$, $p = 0.026$), and signing abilities ($r = .296$, $p = 0.006$); and performance on the MET ($r = .313$, $p = 0.003$) specifically for the melody subtest ($r = .319$, $p = 0.002$) but not for the rhythm subtest ($r = .204$, $p = 0.057$).

Music Exposure

Correlational analysis did not indicate relationship between daily music exposure and discrimination scores ($r = .006$, $p = 0.955$). Daily listening hours and group (between-subject variables) in relation to performance were compared in a two-way ANOVA. Musicians demonstrated improved discrimination of dental-retroflex contrasts with a main effect of group, $F(1, 86) = 9.699$, $p = 0.003$. There was also an interaction effect of group and music listening, $F(1, 86) = 5.429$, $p = 0.022$. While the number of music listening hours a day did not much affect performance in the musician group, there was a significant performance difference among nonmusicians: those who listened to music for at least an hour daily showed better discrimination accuracy than those who listened to less than an hour of music daily.

Instrument Specialization

No statistical differences were found across keyboard, string and wind instrumentalists in the discrimination of dental-retroflex contrasts, $F(2, 43) = 0.621$, $p = 0.543$.

2.3.3b Categorization

Music Training

For categorization of dental-retroflex contrasts, musicians ($M = 92.85$, $SD = 10.09$) significantly outperformed nonmusicians ($M = 83.50$, $SD = 8.79$), $t(31) = 2.816$, $p = 0.008$) as seen in Figure 15. The number of years of music band experience ($r = .492$, $p = 0.004$) and weekly practice hours ($r = .378$, $p = 0.030$) correlated positively with categorization accuracy. However, correlation with years of music training was only marginally significant ($r = .322$, $p = 0.068$).

Figure 15. Mean categorization score across groups. Musicians performed better at categorizing dental-retroflex contrasts than nonmusicians. Error bars represent standard error of the mean (± 1 SE).

Music Aptitude

There was a correlation between categorization scores and general music sophistication scores ($r = .511$, $p = 0.002$) such that participants with higher music sophistication was able to categorize dental-retroflex contrasts more accurately. This correlation was reflected across the five subfactors of sophistication namely, active music engagement ($r = .373$, $p = 0.033$), perceptual abilities ($r = .546$, $p = 0.001$), musical training ($r = .488$, $p = 0.004$), emotions ($r = .435$, $p = 0.011$), and singing

abilities ($r = .428$, $p = 0.013$). Participants in this test were of a different group from those in previous tests and were not evaluated by the MET. As such, correlational data for music aptitude was not available for comparison with performance across groups.

Music Exposure

There was no correlation found for daily music exposure and categorization accuracy ($r = .132$, $p = 0.462$). A two-way ANOVA with group and daily music listening as fixed between-subject factors did not indicate any effect of music exposure on performance, but there was a main effect of group where musicians were better able to categorize dental-retroflex contrasts than nonmusicians, $F(1, 33) = 4.997$, $p = 0.033$.

Instrument Specialization

There were no significant differences between keyboard, string, and wind musicians in categorization scores, $F(2, 13) = 0.078$, $p = 0.926$. This however cannot be construed as an accurate comparison given that only one string and one wind musician were present in the small sample.

3 DISCUSSION

Our present study set out to investigate if positive music-related transfers are possible for segmental contrasts beyond the native language domain. It also sought to explore music experience by looking at different components, namely training, aptitude, exposure, and instrument expertise. In Experiment 1, AX discrimination, short categorization training and ordered discrimination tasks revealed group differences in participants' perceptual sensitivity to nonnative Hindi voicing contrasts, indicating a musician advantage. Experiment 2 also showed that music experience, particularly training, led to better accuracy in discriminating and categorizing nonnative dental-retroflex contrasts. Taken together, the behavioural results of our study support our hypothesis that music experience would result in transfer benefits to nonnative segmental speech perception. We now summarize the main findings of each experiment and discuss possible implications and contributions to the literature.

3.1 Music Experience as a Template for Processing Sounds

Experiment 1 suggests a link between music experience and sensitivity to minute VOT contrasts in nonnative phonemes across AX and ordered discrimination tests. Musicians demonstrated enhanced perception across different sets of unfamiliar voicing contrasts. In addition, it was found that greater music sophistication, aptitude, training, including the number of years of formal training, band experience, and hours of instrument practice a week correlated with better discrimination scores. Although correlation is not causation, the strong statistical relationship between these factors could possibly be used to predict perceptual ability. Moreover, in the ordered discrimination test, which involved higher-level processing skills, auditory attention was found to correlate with task performance across all participants.

An unusual effect was discovered in the exploratory conditions for categorization and ordered discrimination. Participants were primed to activate either speech or nonspeech modes. Musicians performed better when primed to perceive tokens as alien sounds than words of a new language. Conversely, for nonmusicians, discrimination was greatly improved when tokens were perceived as words than alien sounds. This novel finding adds new information to the previous study where all

participants, presumably nonmusicians, demonstrated marked perceptual differences when primed to perceive sine wave analogs as speech or nonspeech (Takayama, 2003). To explain what could be motivating the differences in perception, Takayama posited that while sine wave analogs are functionally equivalent to speech sounds, they do not immediately induce phonemic perception. It is only when listeners are primed to expect speech that they actively attend to acoustical features in the analogs relevant to speech and hence perceive them phonemically. Interestingly, in our study, listener's expectations appeared to interact with music background to override perceptual discrimination. In the case of musicians, perceiving nonnative syllables as nonspeech sounds enhanced the ability to differentiate them. It is possible that tokens were processed analogously to individual music notes, allowing them to pay greater attention to very small duration contrasts in tokens. For nonmusicians, processing unfamiliar syllables as phonemes was more convenient, given that unlike musicians, they lacked a referenceable frame for which to compare tokens as nonspeech sounds. In addition, tokens were naturally-spoken speech syllables, which would have been expected to automatically induce phonemic perception.

3.2 Music Experience as a Facilitator in Language Learning

Experiment 2 explored possible music-related transfers to processing subtle place of articulation differences in dental-retroflex contrasts. We found clear signs of musician advantage in both discrimination and categorization of these contrasts even across different participant samples, e.g. participants in the discrimination test were separate from those in the categorization test. While previous studies have demonstrated some degree of successful learning of nonnative contrasts by means of explicit training (Myers & Swan, 2012; Golestani & Zatorre, 2004), the results are attributed largely to individual learning curves and gradual changes in participants' localization of category boundary due to prolonged training exposure. The findings from our study present new evidence that music experience, specifically training, provides as an additional dimension to accelerate nonnative language learning, where phonetic boundary changes result from both practice/exposure as well as a heightened acuity to shared auditory properties across music and speech.

3.3 The Role of Other Components in Music Experience

Our results suggest that besides music training, music experience may be defined by a composite of factors, viz. aptitude, sophistication, exposure, and instrument expertise, some of which show promise of predicting positive transfers in language processing. There is evidence that aptitude and sophistication both strongly correlate with perceptual ability. For instrument specialization, keyboard musicians in our study outperformed string and wind musicians in discrimination and categorization of VOT contrasts. It is possible that string and wind musicians pay greater attention

to pitch and timbre features given that they devote certain lengths of time to tune their instruments each time before practice in contrast to keyboard musicians. This hypothesized increased sensitivity to timbre could not however be tested in Experiment 2 given the lack of appropriate samples in each instrument category group, e.g. only one string and one wind musician. To note, our findings of differences between instrumentalists should be taken cautiously given that keyboard musicians formed a majority of the musician group and quite a number of musicians reported concurrently playing string, wind, or other percussion instruments as nonprimary instruments.

Our findings also indicate that contrary to prediction, music exposure had no influence on discrimination and categorization of nonnative syllables across all tests. The amount of daily music exposure was a factor included in the Goldsmith Musical Sophistication test, on a 7-point scale ranging from 0.5 hour to > 4 hours. Overall, there was no statistical difference between participants who reported actively listening to more than an hour of music every day and those who reported less than an hour of music daily. This suggests that music exposure may not necessarily result in enhanced perceptual sensitivity for speech.

3.4 Consequences for Theoretical Frameworks

In light of frameworks that explain the music-language relationship, our findings show music training as the most defining factor for influencing perceptual sensitivity to sounds in both music and speech. Music training was shown to enhance perceptual sensitivity to acoustic properties of duration and timbre, which when present in speech sounds, were duly processed with greater efficiency. This is consistent with the OPERA model where salient acoustic features in pitch, timbre and duration are detected by a domain-general sound learning mechanism and in turn similar acoustic information in a different domain is recognized and deemed relevant to effect learning. The greater the overlap of auditory features and neural processing in both domains, the more likely the transfer advantage is posited to occur (McMullen and Saffran, 2004).

While our findings demonstrate music-to-language transfers, a number of studies have emphasized bidirectional effects across music and language domains such that musicianship have enhanced lexical tone perception (Tang, Xiong, Zhang, Dong & Nan 2016; Burnham, Brooker & Reid, 2015; Song, Skoe, Banai & Kraus, 2012; Kraus, Skoe, Parbery-Clark & Ashley, 2009), while a tone language background have likewise contributed to perceptual pitch sensitivity in music (Creel, Weng, Fu, Heyman & Lee, 2018;2017; Bidelman, Hutka & Moreno, 2013; Alexander, Bradlow, Ashley & Wong, 2008).

In our study, two groups of participants having native tone and non-tonal language backgrounds, e.g. English-Chinese and English-Malay bilinguals, were included. However, the effect of music experience was the primary interest, and language background was not a main consideration. Given this, an equal distribution of English-Chinese and English-Malay participants was not ensured. Thus, additive effects of language background and musicianship could not be meaningfully evaluated: that is, we were not able to examine possible language group differences between tone and non-tonal language speakers. However, we will discuss recent results of cross-domain transfers found in the literature.

Apart from well-attested evidence of transfer from music to language, linguistic background have been found to result in positive transfer effects. More specifically, tone languages were shown to enhance perceptual sensitivity to music. Thai (Stevens, Keller & Tyler, 2013) and Cantonese native speakers (Bidelman, Hutka & Moreno, 2013) demonstrated superior pitch attunement in tone memory and melody discrimination than non-tonal language speakers. In another study, Mandarin Chinese nonmusicians showed comparable accuracy to English musicians in categorizing melodic tones (Chang, Hedberg & Wang, 2016). Not surprisingly, tone languages also facilitate perception and learning of linguistic tones. Listeners with native tone language backgrounds demonstrate greater accuracy in discriminating lexical pitch contrasts than non-tonal listeners (Li, C.S.T & Ng, 2017; Tang, Xiong, Zhang, Dong & Nan, 2016), Krishnan, Gandour & Bidelman, 2010; Lee & Lee, 2010; Pfordresher & Brown, 2009), suggesting a direct transfer of pitch processing skill across linguistic and music domains. Furthermore, additive effects of tone language on music training are proposed to contribute to absolute pitch ability (Lee & Lee, 2010; Deutsch, 2002). In a study, (Deutsch, Henthorn, Marvin & Xu, 2006), given a sample of music conservatory students, an approximate 53% of tone language speakers (Chinese) were reported to possess this unique skill compared to 7% of non-tonal language speakers (American).

A number of studies also showed facilitative effects of tone languages in relation to song melody and music perception. Given the fact that each word in tone languages carries lexical pitch, how this pitch is represented in sung melody remains a question of interest. It has been suggested that in some cases, linguistic pitch is often used to determine sung melody, e.g. tones are retained in melody to avoid ambiguity of word meaning. Examples of this can be found across a large number of Mandarin Chinese (Wee, 2007) and Cantonese songs (Yung, 1989). A recent study also evaluated the influence of a tone language background on sung pseudowords containing contrastive pitch in speech not unlike lexical pitch contrasts. Dutch-Cantonese bilinguals (tone language background) and monolingual Dutch (non-tonal language background) judged sung pseudowords in a speeded

task to classify the token in relation to musical and phonological features (Asaridou., Hagoort & McQueen, 2015). To note, while researchers observed a more holistic approach in processing sung pseudowords, there were no observed statistical differences for native tonal language speakers' and nonmusician non-tonal language speakers' performance in discriminating pitch and music intervals. This conclusion may be referenced to a study by Bidelman, Gandour & Krishnan (2011) showing a behavioral and neural performance disparity by musicians and tone language speakers (Chinese). Tone language speakers showed a more robust encoding of musical pitch in brainstem representation that was comparable to encoding in musicians. However, tone language speakers were found to perform similarly to nonmusicians in a behavioral pitch discrimination task. In another study (Hutka, Bidelman & Moreno, 2015) tone language speakers (Cantonese) and musicians both outperformed nonmusicians on pitch discrimination yet only musicians demonstrated enhanced brain response to pitch differences. The results suggest that while there are overlaps in music and language domains, processing differences for perceptual and cognitive transfers likewise exist. In fact, given that music experience afford more extensive training in attending, perceiving and producing a wider range of acoustic features for pitch, timbre, and duration, it could well be more advantageous to auditory acuity than a tone language background.

In relation to the top-down bottom-up framework, music experience is conversely posited to improve both cognitive and perceptual processes, e.g. higher and lower-level functions. To note, in our study, even though short-term learning and categorization of unfamiliar voicing and dental-retroflex contrasts surely involved higher-level cognitive processes, there was no concrete evidence that a musician advantage was motivated by improved higher-level processing, e.g. attention, working memory. No significant group differences were found across group scores in screening tests evaluating these skills. A very recent study (Slater, Azem, Nicol, Swedenborg & Kraus, 2017) reported similar findings. Across groups, voice musicians, percussion musicians and nonmusicians demonstrated no statistical difference in attention, and there were mixed results among musicians for inhibitory control. Taken together, these findings may lead us to possibly rethink the assumption proposed in the literature that music-to-speech transfers are effected indirectly by enhanced higher-level processing mechanisms brought about by music experience instead of resulting from training effects gained through additional learning.

3.5 Future Directions

Our study extends the scope of prior research to explore music-related transfers to phonetic discrimination and categorization beyond the native language domain. There has been a number of reviews positing this possibility by referencing past studies on positive transfers to native contrasts and correlational studies that find music experience improving second-language learning

abilities, such as pronunciation, phonology perception, lexical and syntax learning. Yet there has not been much concrete data to address this hypothesis. Our study provides substantial evidence of positive transfers in this direction. It is also one of very few that report a musician advantage for processing nonnative segmental contrasts, a topic which remains largely unexplored in comparison with a profusion of studies investigating the effect of music experience on processing unfamiliar prosodic speech features, e.g. lexical tone. Building on our work, future studies could investigate in more detail the specific components of music experience which motivate positive transfers at the segmental level with a view to develop an efficient paradigm for nonnative language learning.

3.6 Conclusion

To conclude, our study indicates that music experience as an integrated whole of various components work together to bring out beneficial music-related transfers. There also appears to be an interaction of music training and listener's expectation which in turn influences perception of auditory tokens. In addition, music experience, particularly the component of music training, is a sufficient condition to facilitate improved learning of phonetic contrasts even to include nonnative categories. Positive transfers in our findings support the view of shared auditory features in music and language, and the hypothesis of a domain-general sound learning mechanism used to process sound tokens across domains. Findings are relevant to broaden our understanding of how music experience may be used to effectively to bypass perceptual filters that develop with native language acquisition for second language learning.

REFERENCES

- Abdul-Kareem, I., Stancak, A., Parkes, L., Al-Ameen, M., AlGhamdi, J., Aldhafeeri, et al. (2011). Plasticity of the Superior and Middle Cerebellar Peduncles in Musicians Revealed by Quantitative Analysis of Volume and Number of Streamlines Based on Diffusion Tensor Tractography. *Cerebellum*, 10(3), 611-623.
- Abrams, D., Bhatara, A., & Ryali, S. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex*, 21, 1507-1518.
- Allen, J., Kraus, N., & Bradlow, A. (2000). Neural representation of consciously imperceptible speech sound differences. *Attention, Perception & Psychophysics*, 62, 1383-1393.
- Alexander, J. A., Bradlow, A. R., Ashley, R. D., & Wong, P. C. M. (2008). Music melody perception in tone-language- and nontone-language speakers. *The Journal of the Acoustical Society of America*, 124(4), 2495. doi:10.1121/1.4782815
- Angenstein, N., Scheich, H., & Brechmann, A. (2012). Interaction between bottom-up and top-down effects during the processing of pitch intervals in sequences of spoken and sung syllables. *Neuroimage*, 61(3), 715-722.
- Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology* 83, 111-130.
- Asaridou, S. S., Hagoort, P., & McQueen, J. M. (2015). Effects of Early Bilingual Experience with a Tone and a Non-Tone Language on Speech-Music Integration. *Plos ONE*, 10(9), 1-20.
- Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: Evidence for shared domain-general mechanisms. *Frontiers in Psychology*, 4, 321-321.
- Bao, Z. (2003). Moira Yip (2002). Tone. (Cambridge Textbooks in Linguistics.) Cambridge: Cambridge University Press. pp. xxxiv+341. *Phonology*, 20(2), 275-279.
- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS One*, 8(4), e60676
- 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.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011a) Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, 23(2), 425-434.

- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011b). Musicians demonstrate experience-dependent brainstem enhancement of musical scale features within continuously gliding pitch. *Neuroscience Letters*, 503(3), 203-207.
- Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, 1355, 112-125.
- Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience*, 33(3), 530-538.
- Bidelman, G. M., Schug, J. M., Jennings, S. G., & Bhagat, S. P. (2014). Psychophysical auditory filter estimates reveal sharper cochlear tuning in musicians. *The Journal of the Acoustical Society of America*, 136(1), EL33-EL39.
- Bidelman, G., Weiss, M., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, 40(4), 2662-2673.
- Bidelman, G. M., & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(3), 1240-1249.
- Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex*, 19(7), 1583-1596.
- Blechner, M. J. (1977). *Musical Skill and the Categorical Perception of Harmonic Mode*
- Boh, B., Herholz, S. C., Lappe, C., & Pantev, C. (2011). Processing of Complex Auditory Patterns in Musicians and Nonmusicians. *Plos ONE*, 6(7), 1-10.
- Bradley, E. D. (2016). Phonetic Dimensions of Tone Language Effects on Musical Melody Perception. *Psychomusicology: Music, Mind & Brain*, 26(4), 337-345.
- Brandler, S. (2003). Differences in mental abilities between musicians and non-musicians. *Psychology of Music*, 31(2), 123-138.
- Brown, Steven (2001). The "musilanguage" model of music evolution. In Nils L. Wallin et. alt. (Eds.), *The Origins of Music* (pp. 271-301). Cambridge: MIT Press.
- Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jancke, L. (2005). Scanning silence: Mental imagery of complex sounds. *Neuroimage*, 26, 1119-1127.
- Burns, E. M., & Ward, W. D. (1978). Categorical perception -phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America*, 63(2), 456-468.

- Burnham, D., Brooker, R., & Reid, A. (2015). The effects of absolute pitch ability and musical training on lexical tone perception. *Psychology of Music*, 43(6), 881-897.
- Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature* 396(6707), 128.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, 108, 1-9.
- Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, 25(3-4), 195.
- Chang, D., Hedberg, N., & Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America*, 139(5), 2432.
- Chartrand, J., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, 405(3), 164-167.
- Chao, K.Y., Peng, J.F., Yang, J. C., & Chen, L. (2008). Proceedings from the 2013 IEEE International Symposium on Multimedia. "A Cross-Language Study of Stop Aspiration: English and Mandarin Chinese." Vol. 00, pp. 556-561, Berkeley, CA.
- Chobert, J., Marie, C., Francois, C., Schon, D., & Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *Journal of Cognitive Neuroscience*, 23, 3874-3887.
- Chobert, J., François, C., Velay, J., & 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 (New York, N.Y.: 1991)*, 24(4), 956-967.
- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27, 207-29.
- Coleman, J. (2006). *Introduction to Acoustic Phonetics 4*. [PDF document]. Retrieved from <http://www.phon.ox.ac.uk/jcoleman/AcousticPhonetics4.pdf>
- COMDJ (Artist). (2009). Waveform (amplitude as a function of time) of the English word "above". Retrieved from Wikimedia Commons, <https://commons.wikimedia.org/wiki/File:Waveform-above.png>
- Conference on Music, Language and the Brain, celebrating 25th anniversary of Lerdahl and Jackendoff's Generative Theory of Tonal Music. (2008, January). Prospectus presented at the colloquium, Dijon, France.
- Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *Journal of Acoustical Society of America*, 131, 4756-4769.

- Cooper, A., Wang, Y., & Ashley, R. (2017). Thai rate-varied vowel length perception and the impact of musical experience. *Language and Speech*, 60(1), 65-84.
- Creel, S. C., Weng, M., Fu, G., Heyman, G. D., & Lee, K. (2018; 2017). Speaking a tone language enhances musical pitch perception in 3–5-year-olds. *Developmental Science*, 21(1), n/a. doi:10.1111/desc.12503
- Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., & Frisina, R. D. (1994). Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *Journal of The Acoustical Society of America*, 95(5, Pt 1), 2720-2727.
- Dankovicová, J., House, J., Crooks, A., & Jones, K. (2007). The relationship between musical skills, music training, and intonation analysis skills. *Language and Speech*, 50(2), 177-225.
- Davenport, M., & Hannahs, S. J. (2005). *Introducing Phonetics & Phonology* (2nd ed.). New York; London: Hodder Arnold.
- Dees, T., Russo, N. M., Wong, P. C. M., Kraus, N., & Skoe, E. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4), 420-422.
- Dege, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology*, 2, 124.
- Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research*, 1455, 75-89.
- Delogu, F., Lampis, G., & Olivetti Belardinelli, M. (2006). Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing*, 7, 203-207.
- Deutsch, D. (2002). "The puzzle of absolute pitch," *Curr. Dir. Psychol. Sci.* 11, 200–204.
- Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*, 119(2), 719-722.
- Downing, L. J., & Rialland, A. (2017). Intonation in African Tone Languages (Eds). Volume 24 of Phonological and Phonetics. Berlin, De Gruyter Mouton.
- Duanmu, San. 2000. *The Phonology of Standard Chinese*. Oxford; New York: Oxford University Press.
- Dupoux, E., Peperkamp, S., & Sebastian-Galles, N. (2001). A robust method to study stress "deafness". *Journal of the Acoustical Society of America*, 110(3), 1606-1618.

- Dutta, I. (2007). *Four -way stop contrasts in Hindi: An acoustic study of voicing, fundamental frequency and spectral tilt*
- Dworkis, I. (2012). *The perception of synthesized Swedish vowels: A comparison between musicians and non-musicians*
- Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex*, 49(10), 2812-2821.
- Elmer, S., Klein, C., Kühnis, J., Liem, F., Meyer, M., & Jäncke, L. (2014). Music and language expertise influence the categorization of speech and musical sounds: Behavioral and electrophysiological measurements. *Journal of Cognitive Neuroscience*, 26(10), 2356-2369.
- Elmer, S. (2016). Relationships between music training, neural networks, and speech processing. *International Journal of Psychophysiology*, 108, 46.
- Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex*, 22, 650-658.
- Engineer, N. D., Percaccio, C. R., Pandya, P. K., Moucha, R., Rathbun, D. L., & Kilgard, M. P. (2004). Environmental enrichment improves response strength, threshold, selectivity, and latency of auditory cortex neurons. *Journal of Neurophysiology*, 92, 73-82.
- Escudero, P., & Williams, D. (2014). Distributional learning has immediate and long-lasting effects. *Cognition*, 133408-413.
- Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, 130(4), EL206-EL212. doi:10.1121/1.3629144
- Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Memory and Cognition*, 1, 1-9.
- Fischer -Jørgenson, E. (1954). Acoustic analysis of stop consonants. *Miscellanea Phonetica* 2, 42-59.
- François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, 23(9), 2038-2043.
- Franklin, M. S., Sledge Moore, K., Yip, C., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, 36(3), 353-365.
- Fry, D. B. (1955). Duration and Intensity as Physical Correlates of Linguistic Stress. *Journal of The Acoustical Society Of America*, 27(4), 765.
- Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 23(27), 9240-9245.

- George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, 49(5), 1083-1094.
- Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *Neuroimage*, 21, 494-506.
- Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, 109, 55-67.
- Gordon E. E. (1989). *Advanced Measures of Music Audiation*. Chicago: Riverside Publishing Company.
- Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and mandarin tone discrimination and imitation. *Journal of the Acoustical Society of America*, 115, 2545.
- Gottfried, T. L., & Xu, Y. (2008). Effect of musical experience on mandarin tone and vowel discrimination and imitation. *The Journal of the Acoustical Society of America*, 123(5), 3887-3887.
- Gramley, V. (n.d.) *Articulatory-Acoustic-Auditory Phonetics*. [Powerpoint slides]. Retrieved from <http://www.uni-bielefeld.de/lili/personen/vgramley/teaching/HTHS/review.pdf>
- Gromko, E. J. (2005). The Effect of Music Instruction on Phonemic Awareness in Beginning Readers. *Journal of Research in Music Education*, 53(3), 199-209.
- Güçlü, B., Sevinc, E., & Canbeyli, R. (2011). Duration discrimination by musicians and nonmusicians. *Psychological Reports*, 108, 675-687.
- Hall, J., Owen Van Horne, A., & Farmer, T. (2018). Distributional learning aids linguistic category formation in school-age children. *Journal of Child Language*, 45(3), 717-735.
- Hass, J. (2013). *Chapter One: An Acoustics Primer*. Retrieved from http://www.indiana.edu/~emusic/etext/acoustics/chapter1_timbre.shtml
- Hara, N., Cauvet, E., Devauchelle, A., Le Bihan, D., Dehaene, S., & Pallier, C. (2009). Neural correlates of constituent structure in language and music. *Neuroimage*, 47, S143-S143.
- Hauser, I. (2016). *VOT Variation and Perceptual Distinction* [PDF document]. Retrieved from <http://blogs.umass.edu/ihauser/files/2013/09/lisa2016-slides1.pdf>
- Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press. Retrieved from http://books.google.com.sg/books/about/Metrical_Stress_Theory.html?id=ST1JpDcrR3sC&redir_esc=y
- Herdener, M., Humbel, T., Esposito, F., Habermeyer, B., Cattapan-Ludewig, K., & Seifritz, E. (2014). Jazz drummers recruit language-specific areas for the processing of rhythmic structure. *Cerebral Cortex*, 24(3), 836-843.

- Herholz, S. C., Lappe, C., & Pantev, C. (2009). Looking for a pattern: An MEG study on the abstract mismatch negativity in musicians and nonmusicians. *BMC Neuroscience*, 10(1), 42-42.
- Ho, Y., Cheung, M., & Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology*, 17(3), 439-450.
- Hong, A. Q. (2012). *A phonological and phonetic description of Singapore Hokkien*.
- Hunnicut, L., & Morris, P. A. (2016) "Prevoicing and Aspiration in Southern American English." *University of Pennsylvania Working Papers in Linguistics*: Vol. 22: Is. 1, Article 24. Retrieved from <http://repository.upenn.edu/pwpl/vol22/iss1/24>
- Hutchinson, S., Lee, L. H., Gaab, N., & Schlaug, G. (2003). Cerebellar volume of musicians. *Cerebral Cortex (New York, N.Y.: 1991)*, 13(9), 943-949.
- Hutka, S., Bidelman, G. M., & Moreno, S. (2015). Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia*, 71, 52-63.
- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *The Journal of Neuroscience: The Official Journal of the Society For Neuroscience*, 29(10), 3019-3025.
- Imfeld, A., Oechslin, M. S., Meyer, M., Loenneker, T., & Jancke, L. (2009). White matter plasticity in the corti-cospinal tract of musicians: a diffusion tensor imaging study. *Neuroimage*, 46, 600-607.
- Jacewicz, E., Fox, R. A., & Lyle, S. (2009). Variation in stop consonant voicing in two regional varieties of American English. *Journal of the International Phonetic Association*, 39(3), 313-334.
- Jacques, G. (2011). A panchronic study of aspirated fricatives, with new evidence from pumi. *Lingua*, 121(9), 1518-1538.
- Jain, C., Mohamed, H., & Kumar, U. A. (2015). The effect of short-term musical training on speech perception in noise. *Audiology Research*, 5(1), 5-8.
- Jancke, L., & Shah, N. J. (2004). 'Hearing' syllables by 'seeing' visual stimuli. *European Journal of Neuroscience*, 19, 2603-2608.
- Jones, J. L., Lucker, J., Zalewski, C., Brewer, C., & Drayna, D. (2009). Phonological processing in adults with deficits in musical pitch recognition. *Journal of Communication Disorders*, 42, 226-234.
- Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion (Washington, D.C.)*, 1(4), 381-412.

- Kaganovich, N., Kim, J., Herring, C., Schumaker, J., MacPherson, M., & Weber-Fox, C. (2013). Musicians show general enhancement of complex sound encoding and better inhibition of irrelevant auditory change in music: An ERP study. *European Journal of Neuroscience*, 37(8), 1295-1307.
- Katz, J., Chemla, E., & Pallier, C. (2015). An attentional effect of musical metrical structure. *PLoS One*, 10(11), e0140895.
- Kauramäki, J., Jaaskelainen, I.P., & Sams, M., (2007). Selective attention increases both gain and feature selectivity of the human auditory cortex. *PLoS One* 2, e909.
- Keating, P. A. (1984) Phonetic and phonological representation of stop consonant voicing. *Language*, 60, 286-319.
- Keating, P.A., Linker, W., & Huffman, M. (1983). Patterns in Allophone Distribution for Voiced and Voiceless Stops. *Journal of Phonetics*, 11, 277-290.
- Kempe, V., Bubltz, D., & Brooks, P. J. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology*, 106(2), 349-366.
- Kessinger, R., & Blumstein, S. (1997). Effects of Speaking Rate on Voice-Onset Time in Thai, French, and English. *Journal of Phonetics*, 25, 143-168.
- 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, 125-143.
- Kliuchko, M., Heinonen-Guzejev, M., Monacis, L., Gold, B. P., Heikkilä, K. V., Spinoza, V., Tervaniemi, M., & Brattico, E. (2015). The association of noise sensitivity with music listening, training, and aptitude. *Noise and Health*, 17(78), 350-357
- Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Frontiers in Psychology*, 2, 110, 1-20.
- Koelsch, S., Kasper, E., Sammler, D., Schulze K., Gunter, T., & Friederici A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, 7, 302-307.
- Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between Syntax Processing in Language and in Music: An ERP Study. *Journal of Cognitive Neuroscience*, 17(10), 1565-1577.
- Kohler, K. J. (1979). Dimensions in the perception of fortis and lenis plosives. *Phonetica*, 36(4-5), 332-343.
- Kolinsky, R., Cuvelier, H., Goetry, V., Peretz, I., & Morais, J. (2009). Music Training Facilitates Lexical Stress Processing. *Music Perception: An Interdisciplinary Journal*, 3, 235-246.

- Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery - Sound of silence activates auditory cortex. *Nature*, 434, 158.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599-605.
- 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, 543 -557.
- Kraus, N., & Slater, J. (2015). Music and language: relations and disconnections In Celesia, G.C. & Hickok, G. (Eds.), *Handbook of Clinical Neurology*, Vol. 129 (3rd series) *The Human Auditory System*. Elsevier.
- Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, 126(1), 159-179.
- Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and Language Syntax Interact in Broca's Area: An fMRI Study. *Plos One*, 10(11), e0141069.
- Kühnis, J., Elmer, S., Meyer, M., & Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia*, 51(8), 1608-1618.
- Lappe, C., Trainor, L. J., Herholz, S. C., & Pantev, C. (2011). Cortical Plasticity Induced by Short -Term Multimodal Musical Rhythm Training. *Plos ONE*, 6(6), 1-8.
- Lee, C., Lekich, A., & Zhang, Y. (2014). Perception of pitch height in lexical and musical tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, 135(3), 1607-1615.
- Lee, C. H., & Hung, T. H. (2008). Identification of Mandarin tones by English speaking musicians and non-musicians. *Journal of the Acoustical Society of America* 124, 3235-3248
- Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *Journal of Neuroscience*, 29(18), 5832-5840.
- Lee, C., & Lee, Y. (2010). Perception of musical pitch and lexical tones by Mandarin-speaking musicians. *The Journal of the Acoustical Society of America*, 127(1), 481-490.
- Levitin, D. J., & Menon, V. (2003). Regular article: Musical structure is processed in “language” areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *Neuroimage*, 20, 2142-2152.
- Li, B., Oh, S., Shao, J., & Shuai, L. (2012). Reciprocal perception of Chinese and Korean affricates and fricatives. *The Journal of the Acoustical Society of America*, 131(4), 3272.
- Li, X., C.S.T., & Ng, M. L. (2017). Effects of L1 tone on perception of L2 tone - a study of mandarin tone learning by native Cantonese children. *Bilingualism*, 20(3), 549. doi:10.1017/S1366728916000195

- Lim, L. (2010). Peranakan English in Singapore In Schreier, D., Trudgill, P., Schneider, E. W., & Williams, J. P (Ed.), *The lesser-known varieties of English: An introduction* (pp. 338-342). Cambridge: Cambridge University Press.
- Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, 11, 1021-1031.
- Lisker, L., & Abramson, A. (1964). "A Cross-language Study of Voicing in Initial Stops". *Word*, 20, 384-422.
- Locke, S., & Kellar, L. (1973). Categorical Perception in a Non-Linguistic Mode. *Cortex*, 9(4), 355-369.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, N.J: Lawrence Erlbaum Associates.
- Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, 18(2), 199-211.
- Magne, C., Jordan, D. K., & Gordon, R. L. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain and Language*, 153-154, 13-19.
- Mannell, R. (2009, August 1). *Phonetics and Phonology: Voice Onset Time*. Retrieved from http://clas.mq.edu.au/speech/phonetics/phonetics/airstream_laryngeal/vot.html
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in mandarin chinese. *Journal of Cognitive Neuroscience*, 23(10), 2701-2715.
- Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, 23(2), 294-305.
- Mathias S, O., Adrian, I., Thomas, L., Martin, M., & Lutz, J. (2010). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. *Frontiers in Human Neuroscience*, Vol 3 (2010), doi:10.3389/neuro.09.076.2009/full.
- Martínez-Montes, E., Hernández-Pérez, H., Chobert, J., Morgado-Rodríguez, L., Suárez-Murias, C., Valdés-Sosa, P. A., & Besson, M. (2013). Musical expertise and foreign speech perception. *Frontiers in Systems Neuroscience*, 7, 84.
- Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, 19(9), 1453-1463.
- Masataka, N. (2009). Review: The origins of language and the evolution of music: A comparative perspective. *Physics of Life Reviews*, 6, 11 -22.

- McMullen, E., & Saffran, J. R. (2004). Music and Language: A Developmental Comparison. *Music Perception: An Interdisciplinary Journal*, (3), 289.
- 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.
- Milovanov, R., Huotilainen, M., Valimaki, V., Esquef, P.A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. *Brain Research*, 1194, 81 -89.
- Mithen, S., Morley, I., Wray, A., Tallerman, M., & Gamble, C. (2006). The singing neanderthals: The origins of music, language, mind and body, by Steven Mithen. London: Weidenfeld & Nicholson, 2005. ISBN 0-297-64317-7 hardback £20 & US\$25.2; ix+374 pp. *Cambridge Archaeological Journal*, 16(1), 97-112.
- 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, 712-723.
- Moosmüller, S., & Ringen, C. (2004). Voice and aspiration in Austrian German plosives. *Folia Linguistica: Acta Societatis Linguisticae Europaeae*, 38(1-2), 43-62.
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-musicians: An Index for Assessing Musical Sophistication in the General Population. *Plos ONE*, 9(2), 1-23.
- 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 of the United States of America*, 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.
- Myers, E. B., & Swan, K. (2012). Effects of Category Learning on Neural Sensitivity to Non-native Phonetic Categories. *Journal of Cognitive Neuroscience*, 24(8), 1695-1708.
- Nespor, M., Shukla, M., & Mehler, J. (2011). Stress-timed vs. Syllable-timed Languages in van Oostendorp, M., Ewen, C. J., Hume, E. & Rice, K. (Eds.), *The Blackwell Companion to Phonology* (pp. 1147-1159). Malden: Wiley-Blackwell.
- Ng, S. (2005). *Method in the madness? VOT in Singaporean native languages and English*
- Nicholson, K. G., Baum, S., Kilgour, A., Koh, C. K., Munhall, K. G., & Cuddy, L. L. (2003). Impaired processing of prosodic and musical patterns after right hemisphere damage. *Brain and Cognition*, 52(3), 382-389.

- Oldfield, R. C. (1971). The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsychologia*, 9(1), 97 -113.
- Ong, J. H., Burnham, D., Stevens, C. J., & Escudero, P. (2016). Naïve learners show cross-domain transfer after distributional learning: The case of lexical and musical pitch. *Frontiers in Psychology*, 7, 1189. doi:10.3389/fpsyg.2016.01189
- Pantev, C., & Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neuroscience and Biobehavioral Reviews*, 35(10), 2140-2154.
- Pantev, C., Wollbrink, A., Roberts, L. E., Engelien, A., & Lütkenhöner, B. (1999). Short-term plasticity of the human auditory cortex. *Brain Research*, 842(1), 192-199.
- 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., Skoe, E., Lam, C., & Kraus, N. (2009a). Musician enhancement for speech-in-noise. *Ear and Hearing*, 30(6), 653-661.
- Parbery-Clark, A., Skoe, E., & Kraus, N. (2009b). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience*, 29, 14100-14107.
- Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience*, 219, 111-119.
- Park, M., Gutyrchik, E., Welker, L., Carl, P., Pöppel, E., Zaytseva, Y., Meindl, T., Blautzik, J., Reiser, M., & Bao, Y. (2015). Sadness is unique: Neural processing of emotions in speech prosody in musicians and non-musicians. *Frontiers in Human Neuroscience*, 8, doi:10.3389/fnhum.2014.01049.
- Patel, A. D. (2014; 2013). Can nonlinguistic musical training change the way the brain processes speech? the expanded OPERA hypothesis. *Hearing Research*, 308, 98.
- Patel, A. D. (2012). The OPERA hypothesis: assumptions and clarifications. *Annals of the New York Academy of Sciences*, 1252, 124 -128.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? the OPERA hypothesis. *Frontiers in Psychology*, 2, 142.
- Patel, A. D. (2008) *Music, Language, and the Brain*. New York: Oxford University Press.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing Syntactic Relations in Language and Music: An Event-Related Potential Study. *Journal of Cognitive Neuroscience*, 10(6), 717-733.

- Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language*, 61(1), 123-144.
- Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception & Psychophysics*, 71(6), 1385-1398.
- Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, 693-703.
- Peretz, I., Vuvan, D., Lagrois, M., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 370(1664), 20140090.
- Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., & Dilley, L. C. (2013). Evidence for shared cognitive processing of pitch in music and language. *PloS One*, 8(8), e73372.
- Perrot, X., Micheyl, C., Khalfa, S., & Collet, L. (1999). Stronger bilateral efferent influences on cochlear biomechanical activity in musicians than in nonmusicians. *Neuroscience Letters*, 262, 167-170.
- Platt, J., & Weber, H. (1980). *English in Singapore and Malaysia: Status, Features, Functions*. KL: Oxford University Press.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, 89(6), 2961-2977.
- Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, 119(3), 1684-1696.
- Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, 24, 37-48.
- Reinke, K. S., He, Y., Wang, C., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Brain Research, Cognitive Brain Research*, 17, 781-791.
- Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain Potentials to Native and Non-Native Speech Contrasts in 7- and 11-Month-Old American Infants. *Developmental Science*, 8(2), 162-172.
- Rivera-Gaxiola, M., Csibra, G., Johnson, M., & Karmiloff-Smith, A. (2000). Research report: Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research*, 111, 13-23.
- Ross, E. D. (1981). The aprosodias: Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology*, 38(9), 561-569.

- Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, 156(1), 95-103.
- Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross -linguistic study. *Acta Psychologica*, 138(1), 1-10.
- Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995a). Increased corpus callosum size in musicians. *Neuropsychologia*, 33, 1047-1055.
- Schmithorst, V. J., & Wilke, M. (2002). Differences in white matter architecture between musicians and non- musicians: a diffusion tensor imaging study. *Neuroscience Letters*, 321, 57-60.
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, 5, 688-694.
- Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *Neuroimage*, 51(1), 450-461.
- 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.
- Schulze, K., Mueller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience*, 33(1), 189-196.
- Schulze, K., Zysset, S., Mueller, K., Friederici, A.D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in non-musicians and musicians. *Human Brain Mapping*, 32, 771-783.
- Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., & Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3), 743-751.
- Seppänen, M., Hämäläinen, J., Pesonen, A., & Tervaniemi, M. (2013). Passive sound exposure induces rapid perceptual learning in musicians: Event-related potential evidence. *Biological Psychology*, 94(2), 341-353.
- Settari, O. (1997). The Aesthetic Views of Music of Descartes and Comenius. *Musicologica Brunensia*, 46, 5-14.
- Shahidi, A. H., & Aman, R. (2011). An Acoustical Study of English Plosives in Word Initial Position produced by Malays. *3L: Southeast Asian Journal of English Language Studies*, 17(2), 23-33.
- Shahin, A. J. (2011). Neurophysiological influence of musical training on speech perception. *Frontiers in Psychology*, 2, 126.

- Shin, S. J. (2001). Cross-language Speech Perception in Adults: Discrimination of Korean Voiceless Stops by English Speakers. *Studies in Linguistic Sciences, Volume 31, 2*, 155-166.
- Shook, A., Marian, V., Bartolotti, J., & Schroeder, S. (2013). *Musical experience influences novel language learning. American Journal of Psychology, 126*, 95-104.
- Skoe, E., & Kraus, N. (2012) A little goes a long way: how the adult brain is shaped by musical training in childhood. *Journal of Neuroscience, 32*, 11507-11510.
- Skoe, E., Krizman, J., Spitzer, E., & Kraus, N. (2013). The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience, 243*, 104-114.
- Slater, J., Azem, A., Nicol, T., Swedenborg, B., & Kraus, N. (2017). Variations on the theme of musical expertise: Cognitive and sensory processing in percussionists, vocalists and non-musicians. *European Journal of Neuroscience, 45*(7), 952-963.
- Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychological Science, 17*, 675-681.
- Slevc, L. R., & Okada, B. M. (2015). Processing structure in language and music: A case for shared reliance on cognitive control. *Psychonomic Bulletin & Review, 22*(3), 637-652.
- Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to improve hearing speech in noise: Biological mechanisms. *Cerebral Cortex, 22*(5), 1180-1190.
- Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency discrimination tasks by musicians and nonmusicians. *Journal of the Acoustical Society of America, 76*, 1690-1695.
- Stevens, C. J., Keller, P. E., & Tyler, M. D. (2013). Tonal Language Background and Detecting Pitch Contour in Spoken and Musical Items. *Psychology of Music, 41*(1), 59-74.
- Strait, D. L., Hornickel, J., & Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and Brain Functions: BBF, 7*(1), 44-44.
- Strait, D. L., & Kraus, N. (2011). Can you hear me now? musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in Psychology, 2*, 113.
- Strait, D., & Kraus, N. (2011). Playing music for a smarter ear: Cognitive, perceptual and neurobiological evidence. *Music Perception: An Interdisciplinary Journal, 29*(2), 133-146.
- 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.
- Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language, 123*(3), 191-201.

- Suárez, L., Elangovan, S., & Au, A. (2016). Cross-sectional study on the relationship between music training and working memory in adults: Music and working memory. *Australian Journal of Psychology*, 68(1), 38-46.
- Takayama, T. (2004). Priming effects in speech and nonspeech modes of perception. *Acoustical Science and Technology*, 25(3), 196-202.
- Tang, W., Xiong, W., Zhang, Y., Dong, Q., & Nan, Y. (2016). Musical experience facilitates lexical tone processing among Mandarin speakers: Behavioral and neural evidence. *Neuropsychologia*, 91, 247-253.
- Tay, M. W. J. (1993). *The English Language in Singapore: Issues and Developments*. Singapore: Unipress.
- Tee, C. T. C. (1986). *Aspiration in Singapore English: An Instrumental Study*.
- Tees, R. C., & Werker, Janet F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, 38, 579-590.
- Teo, G. A. (2013). *Learning the Hindi dental -retroflex contrast through sound -to -meaning associations and sound discrimination tasks*.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch-discrimination accuracy in musicians vs. non-musicians, An event-related potential and behavioral study. *Experimental Brain Research*, 161, 1-10.
- Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., & Friederici, A. D. (2009). Top-down modulation of auditory processing: Effects of sound context, musical expertise and attentional focus. *The European Journal of Neuroscience*, 30(8), 1636-1642.
- Thomas, D. A., & American Council of Learned Societies. (1995). *Music and the origins of language: Theories from the French enlightenment*. New York; Cambridge: Cambridge University Press.
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion*, 4, 46-64.
- Trainor, L. J., Marie, C., Gerry, D., Whiskin, E., & Unrau, A. (2012). Becoming musically enculturated: Effects of music classes for infants on brain and behavior. *Annals of the New York Academy of Sciences*, 1252(1), 129-138.
- Tsui, I., & Ciocca, V. (2000). Perception of aspiration and place of articulation of Cantonese initial stops by normal and sensorineural hearing-impaired listeners. *International Journal Of Language & Communication Disorders*, 35(4), 507-525.
- Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: Mechanisms of plasticity in the auditory brainstem. *Neuron*, 62(4), 463-469.

- van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research*, 23(2), 270-276.
- Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influences of place of articulation and speaking rate on the internal structure of voicing contrasts. *Journal of the Acoustical Society of America*, 92, 735.
- Wallentin, M., Nielsen, A. H., Friis -Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, 20(3), 188-196.
- Wang, X., Wang, M., & Chen, L. (2013). Hemispheric lateralization for early auditory processing of lexical tones: Dependence on pitch level and pitch contour. *Neuropsychologia*, 51(11), 2238-2244.
- Wee, Lian-Hee. (2007). Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics*. 35. 128-144.
- Weidema, J. L., Roncaglia-Denissen, M. P., & Honing, H. (2016). Top-Down Modulation on the Perception and Categorization of Identical Pitch Contours in Speech and Music. *Frontiers in Psychology*, Vol 7 (2016). Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4889578/>.
- Werker J. F., & Tees, R.C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75, 1866-78.
- Wong, P. C. M., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping*, 28, 995-1006.
- Wong, F. C. K., Chandrasekaran, B., Garibaldi, K., & Wong, Patrick C. M. (2011). White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *The Journal of Neuroscience*, 31 (24), 8780-8785.
- Yoo, S. S., Lee, C. U., & Choi, B. G. (2001). Human brain mapping of auditory imagery: event-related functional MRI study. *Neuroreport*, 12, 3045-3049.
- Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional Phonetic Learning at 10 Months of Age. *Infancy*, 15(4), 420-433.
- Yung, B. (1989). *Cantonese opera: Performance as creative process*. New York;Cambridge [Cambridgeshire]: Cambridge University Press.
- Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science*, (6158), 585-589.

- Zatorre, R. J., & Schönwiesner, M. (2011). "Cortical speech and music processes revealed by functional neuroimaging," in *The Auditory Cortex*, eds. Winer J. A., Schreiner C. E., editors. (Boston, MA: Springer), 657-677.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Review: Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6, 37-46.
- Zendel, B.R., & Alain, C. (2009) Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, 21, 1488-1498.
- Zhao, T., & Kuhl, P. K. (2013). Effects of musical rhythm training on infants' neural processing of temporal information in music and speech. *The Journal of the Acoustical Society of America*, 134(5), 4236-4236.
- Zhu, J., & Chen, Y. (2016). Effect of several acoustic cues on perceiving mandarin retroflex affricates and fricatives in continuous speech. *The Journal of the Acoustical Society of America*, 140(1), 461-470.
- Zhu, L., Xia, J., & Shinn-Cunningham, B. (2011). Relationship between selective auditory attention and brainstem encoding in musicians and non-musicians. *The Journal of the Acoustical Society of America*, 129(4), 2490-2490.
- Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D. E., Tallal, P., & Gaab, N. (2013). Enhanced syllable discrimination thresholds in musicians. *PloS One*, 8(12), e80546.

Appendix A: Participant Language Background

AX discrimination, ordered discrimination and categorization of VOT contrasts

Participant	Languages Known
10358	Chinese, English
10364	Chinese, English, Japanese
10365	English, Mandarin
10357	English, Chinese
10368	English, Chinese, Hokkien, Cantonese
10374	Chinese, English
10386	Chinese, English
10388	English, Chinese, Malay
10396	English, Mandarin, Japanese
10401	English, Chinese
10402	English, Chinese
10399	English, Chinese, Spanish
10378	Mandarin, English, Hokkien, Hainanese
10406	English, Mandarin, Hokkien
10407	English, Chinese, Teochew, Korean, French
10397	English, Chinese, Malay
10409	English, Mandarin, Cantonese
10410	Chinese, English
10411	English, Chinese
10414	English, Cantonese, Japanese, French
10415	English, Chinese
10416	English, Mandarin, Cantonese, learned German
10417	English, Chinese, Korean 1?
10418	English, Malay
10420	English Mandarin
10421	English, Chinese
10422	English, Chinese
10425	English, Chinese, Korean
10428	English, Chinese, Japanese (Beginner)
10431	Indonesian, English, Chinese (Elementary), Japanese (Elementary)
10432	English, Mandarin, Cantonese, Japanese
10433	English, Chinese, French, Japanese

10435	English, Vietnamese, Chinese
10437	English, Vietnamese, Chinese
10438	English, Mandarin, Hokkien, Japanese, Spanish, Swedish
10439	English, Chinese, Shanghai dialect, learning German
10442	English, Mandarin, Japanese
10444	English, Chinese, Malay
10445	English, Sundanese, Indonesian, little bit of Chinese/Korean
10446	English, Mandarin
10448	English, Mandarin, some Cantonese
10449	English Chinese
10450	English, Chinese
10451	English, Chinese
10452	English, Chinese
10453	English, Chinese
10455	English, Mandarin Chinese, Hokkien
10456	Hokkien, Chinese, English, Malay
10457	Mandarin, English, Malay, Korean(Basic)
10459	Mandarin, English, Malay, Hokkien, Cantonese
10462	English, Mandarin Chinese, French
10463	English, Mandarin, Hokkien
10464	English Chinese Malay Cantonese, Hakka
10465	English, Mandarin, Malay
10461	English, Indonesian
10467	Chinese, Malay, English, Hokkien
10468	Hokkien, Cantonese, Malay, Chinese
10469	English, Malay, Korean
10472	Chinese, English, Malay
10474	English Chinese Hokkien
10475	English, Chinese, Malay
10476	English, Mandarin, Japanese
10477	Chinese, English
10478	English, a little Mandarin
10479	English, Chinese, Teochew, Indonesian
10480	English, Chinese, Korean, a bit of Hokkien
10481	English, Chinese
10482	Chinese, English
10484	English, Malay
10485	English, Chinese, Korean
10486	Mandarin, Malay, English

10487	Chinese, English
10488	English, Mandarin, Hokkien
10489	English, Chinese
10490	English, Chinese
10491	Chinese, English
10492	English, Chinese
10496	English, Chinese
10497	Mandarin, English
10496	English, Chinese
10497	Mandarin, English
10494	English Mandarin Japanese
10496	English, Chinese
10497	Mandarin, English
10498	Chinese English
10499	English, Mandarin
10500	English, Chinese, Malay
10501	English, Chinese, Hokkien

AX discrimination and categorization of dental-retroflex contrasts

Participant	Known Languages
10445	English, Sudanese, Indonesian, a little bit Chinese/Korean
10448	English, Mandarin, some Cantonese
10450	English, Chinese
10453	English, Chinese
10455	English, Mandarin Chinese, Hokkien
10456	Hokkien, Chinese, English, Malay
10457	Mandarin, English, Malay, Korean (basic)
10461	Bahasa Indonesia, English
10464	English, Chinese, Malay, Cantonese, Hakka
10465	English, Mandarin, Malay
10474	English Chinese Hokkien
10467	Chinese, Malay, English, Hokkien
10468	Hokkien, Cantonese, Malay, Chinese
10480	English, Chinese, Korean, a bit of Hokkien
10469	English, Malay, Korean
10459	Mandarin, English, Malay, Hokkien, Cantonese
10484	English, Malay
10485	English, Chinese, Korean
10442	English, Mandarin, Japanese
10481	English, Chinese
10477	Chinese, English
10488	English, Mandarin, Hokkien
10487	Chinese, English
10492	English, Chinese, Hokkien
10486	Mandarin, Malay, English
10482	Chinese, English
10493	Chinese, Malay, English
10490	English, Chinese
10491	Chinese, English
10494	English, Mandarin, Japanese
10495	English, Chinese, Hokkien
10496	English, Chinese
10500	English, Chinese, Malay
10503	English, Indonesian, Malay, Chinese

Appendix B: Stimuli

Carrier Sentence: दोबारा _____ एक बोली

/dɔbara/ _____ /ek bɔlo/

Dental Contrasts

1	Voiceless dental plosive	/ta te to/	ता ते तै
2	Voiced dental plosive	/da de do/	दा दे दै
3	Voiceless aspirated dental plosive	/tʰa tʰe tʰo/	था थे थै
4	Voiced aspirated dental plosive	/dʰa dʰe dʰo/	धा धे धै

Retroflex Contrasts

1	Voiceless retroflex plosive	/ʈa ʈe ʈo/	टा टे टै
2	Voiced retroflex plosive	/ɖa ɖe ɖo/	डा डे डै
3	Voiceless aspirated retroflex plosive	/ʈʰa ʈʰe ʈʰo/	ठा ठे ठै
4	Voiced aspirated retroflex plosive	/ɖʰa ɖʰe ɖʰo/	ढा ढे ढै

Velar Contrasts

1	Voiceless velar plosive	/ka ke ko/	का के कै
2	Voiced velar plosive	/ga ge go/	गा गे गै
3	Voiceless aspirated velar plosive	/kʰa kʰe kʰo/	खा खे खै
4	Voiced aspirated velar plosive	/gʰa gʰe gʰo/	घा घे घै

Palatal Contrasts

1	Voiceless palatal plosive	/cɛa cɛe cɛo/	चा चे चै
2	Voiced palatal plosive	/jɛa jɛe jɛo/	जा जे जै
3	Voiceless aspirated palatal plosive	/cɛʰa cɛʰe cɛʰo/	छा छे छै
4	Voiced aspirated palatal plosive	/jɛʰa jɛʰe jɛʰo/	झा झे झै

Appendix C: The Goldsmiths Musical Sophistication Form, v1.0

Please circle the most appropriate category:	1 Completely Disagree	2 Strongly Disagree	3 Disagree	4 Neither Agree nor Disagree	5 Agree	6 Strongly Agree	7 Completely Agree
1. I spend a lot of my free time doing music-related activities.	1	2	3	4	5	6	7
2. I sometimes choose music that can trigger shivers down my spine.	1	2	3	4	5	6	7
3. I enjoy writing about music, for example on blogs and forums.	1	2	3	4	5	6	7
4. If somebody starts singing a song I don't know, I can usually join in.	1	2	3	4	5	6	7
5. I am able to judge if someone is a good singer or not.	1	2	3	4	5	6	7
6. I usually know when I'm hearing a song for the first time.	1	2	3	4	5	6	7
7. I can sing or play music from memory.	1	2	3	4	5	6	7
8. I am intrigued by musical styles I'm not familiar with and want to find out more.	1	2	3	4	5	6	7
9. Pieces of music rarely evoke emotions for me.	1	2	3	4	5	6	7
10. I am able to hit the right notes when I sing along with a recording.	1	2	3	4	5	6	7
11. I find it difficult to spot mistakes in a performance of a song even if I know the tune.	1	2	3	4	5	6	7
12. I can compare and discuss differences between two performances or versions of the same piece of music.	1	2	3	4	5	6	7
13. I have trouble recognizing a familiar song when played in a different way or by a different performer.	1	2	3	4	5	6	7
14. I have never been complimented for my talents as a musical performer.	1	2	3	4	5	6	7
15. I often read or search the internet for things related to music.	1	2	3	4	5	6	7

Please circle the most appropriate category:	1 Completely Disagree	2 Strongly Disagree	3 Disagree	4 Neither Agree nor Disagree	5 Agree	6 Strongly Agree	7 Completely Agree
16. I often pick certain music to motivate or excite me.	1	2	3	4	5	6	7
17. I am not able to sing in harmony when somebody is singing a familiar tune.	1	2	3	4	5	6	7
18. I can tell when people sing or play out of time with the beat.	1	2	3	4	5	6	7
19. I am able to identify what is special about a given musical piece.	1	2	3	4	5	6	7
20. I am able to talk about the emotions that a piece of piece evokes for me.	1	2	3	4	5	6	7
21. I don't spend much of my disposable income on music.	1	2	3	4	5	6	7
22. I can tell when people sing or play out of tune.	1	2	3	4	5	6	7
23. When I sing, I have no idea whether I'm in tune or not.	1	2	3	4	5	6	7
24. Music is a kind of an addiction for me –I couldn't live without it.	1	2	3	4	5	6	7
25. I don't like singing in public because I'm afraid that I would sing wrong notes.	1	2	3	4	5	6	7
26. When I hear a music I can usually identify its genre.	1	2	3	4	5	6	7
27. I would not consider myself a musician	1	2	3	4	5	6	7
28. I keep track of new music that I come across (e.g. new artists or recordings).	1	2	3	4	5	6	7
29. After hearing a new song two or three times, I can usually sing it by myself.	1	2	3	4	5	6	7
30. I only need to hear a new tune once and I can sing it back hours later.	1	2	3	4	5	6	7
31. Music can evoke my memories of past people and places.	1	2	3	4	5	6	7
32. I engage in regular daily practice of a musical instrument (including voice) for ____ years.	0	1	2	3	4-5	6-9	>10

Please circle the most appropriate category:	1 Completely Disagree	2 Strongly Disagree	3 Disagree	4 Neither Agree nor Disagree	5 Agree	6 Strongly Agree	7 Completely Agree
33. At the peak of my interest, I practiced ____ hours per day on my primary instrument.	0	0.5	1	1.5	2	3-4	≥ 5
34. I have attended ____ live music events as an audience in the past twelve months.	0	1	2	3	4-6	7-10	≥ 11
35. I have had formal training in music theory for ____ years.	0	0.5	1	2	3	4-6	≥ 7
36. I have had ____ of formal training on a musical instrument (including voice) during my lifetime.	1	2	3	4	5	6	7
37. I can play ____ musical instruments.	0	0.5	1	2	3-5	6-9	≥ 10
38. I listen intently to music for ____ per day.	0-15 min	15-30 min	30-60 min	60-90 min	2 hrs	2-3 hrs	≥ 4 hrs
39. The instrument I play best (including voice) is ____							