MECHANISMS OF PITCH DISCRIMINATION IN MUSICIANS AND NON-MUSICIANS

LAUREN H. VOMBERG
Bachelor of Arts and Science, University of Lethbridge, 2017

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

MASTER OF SCIENCE

in

PSYCHOLOGY

Department of Psychology
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

MECHANISMS OF PITCH DISCRIMINATION IN MUSICIANS AND NON-MUSICIANS

LAUREN H. VOMBERG

Date of Defence: April 26, 2019

Dr. J. R. Vokey	Professor	Ph.D.
Dr. S. W. Allen	Associate Professor	Ph.D.
Thesis Co-Supervisors		
Dr. J. Sadr	Associate Professor	Ph.D.
Thesis Examination Committee Member		
Dr. B. Hughes	Assistant Professor	Ph.D.
Thesis Examination Committee Member		
Dr. J. Williams	Instructor	Ph.D.
Chair, Thesis Examination Committee		

DEDICATION

To my parents, for all their love, support, and patience.

ABSTRACT

Musicians are better able to discriminate changes in pitch than non-musicians, particularly when judging tones constructed without the fundamental frequency, and especially when asked to vocally reproduce the tones. In two experiments, musicians and non-musicians were asked to judge whether a test tone was higher or lower than a reference tone. Participants were tested across four conditions, three designed to directly control humming; no specific instructions regarding humming specifically asked to hum, and speeded response (no time to hum). In the fourth condition, participants responded by moving a vertical slider up or down on the computer screen to the extent that they thought the tones differed. In Experiment 1, musicians were more accurate in performance accuracy and vocal reproduction. In Experiment 2, there were no statistically significant differences between participants.

ACKNOWLEDGMENTS

I would like to express my very great appreciation to my supervisors, Dr. John Vokey and Dr. Scott Allen, for their patient guidance, incredible encouragement, and useful critiques throughout the implementation of this thesis. I would also like to thank Dr. Javid Sadr and Dr. Bryn Hughes for their support and guidance, as well as Dr. Janet Youngdahl for her assistance in regard to vocal pedagogy.

I would also like to extend my gratitude to Mrs. Leanne Wehlage-Ellis for her administrative support.

Thank you to my parents, Mac and Janice Vomberg, for always encouraging my love of music and academia. I would also like to extend my deepest gratitude to my aunt, Elisabeth Shapiro, for all of her support throughout this academic journey.

Last, and finally not least, the members of the Vallen lab. It has been an absolute joy to share the lab space with you all for the last two years, and I could not have completed this thesis without all of you.

TABLE OF CONTENTS

Dedication……………………………………………………………………..………iii

Abstract…………………………………………………………………....………..iv

Acknowledgements………………………………………………………..…………v

List of Tables…………………………………………………………………………..vi

List of Figures………………………………………………………...……………vii

Chapter 1: Introduction…..……………………………………………………………...1

1.1 Sound…………………………………………………………………………..2

1.1.1 Pitch………………………………………………………………..2

1.1.2 Auditory Scene Analysis………………………………………..4

1.1.3 Frequency…………………………………………………………6

1.1.4 Fundamental Frequency………………………………………...6

1.1.5 Harmonic Analysis……………………………………………...7

1.1.6 Overtone Series…………………………………………………9

1.1.7 The Phenomenon of the Missing Fundamental……………………9

1.2 Pitch Perception…………………………………………………………10

1.2.1 Stimuli in Pitch Perception Research……………………………13

1.2.2 Individual Differences…………………………………………...13

1.2.3 Pitch Reproduction………………………………………………16

1.2.4 Tonal Memory……………………………………………………20

1.2.5 Vocal Range……………………………………………………...20

1.3 Influence of Language……………………………………………………..22

1.4 Previous Research…………………………………………………………24

Chapter 2: The phenomenon of the missing fundamental: Seither-Preisler et al. (2007)...26

2.1 Participants……………………………………………………………………...26

2.11 Professionals, Non-Musicians, and Musical Amateurs……………..26

2.2 The Auditory Ambiguity Test (AAT)………………………………….....27

2.2.1 Spectral Construction of the Tones………………………………...28

2.2.2 Procedure……………………………………………………….......28

2.3 Data Analysis…………………………………………………………….....29

2.4 Results…………………………………………………………………….29

2.4.1 Effect of Musical Competence……………………………………..29

2.4.2 Effect of Interval Width……………………………………………...30

2.5 Discussion...33

2.5.1 Hypothesis 1...34

2.5.2 Hypothesis 2...35

2.5.3 Hypothesis 3...35

2.6 Conclusion...35

2.7 Critical Analysis of Seither-Preisler et al. (2007)...36

Chapter 3: Research following Seither-Preisler et al. (2007)..39

3.1 Replicating Seither-Preisler et al. (2007): Granzow (2010)............................39

3.2 A step further: Reimchen (2011)...47

Chapter 4: Overview of the Experiments...50

4.1 Experimental Design...50

4.2 Conditions...53

4.2.1 No Instruction...53

4.2.2 Vocal Reproduction...53

4.2.3 Slider...54

        4.2.4 Speed..54

    4.3 Participants..54

        4.3.1 Modified Latin Square Design..56

    4.4 Overtones and Frequencies...57

    4.5 Spectral Series..57

        4.5.1 Low Tones...59

        4.5.2 Middle Tones...59

        4.5.3 High Tones...59

Chapter 5: Experiment 1: Ramped Tones..63

    5.1 Participants..63

    5.2 Materials and Procedure..63

    5.3 Results..64

        5.3.1 Proportion Correct...65

        5.3.2 Vocal Reproduction...66

        5.3.3 Key-Press versus Sung Direction..66

        5.3.4 Sung Frequencies of the Reference Tones..68

        5.3.5 Sung Frequencies of the Test Tones..69

    5.4 Discussion...69

Chapter 6: Experiment 2: Non-Ramped Tones..78

    6.1 Participants..78

    6.2 Materials and procedure...79

    6.3 Results..79

        6.3.1 Proportional Correct..80

    6.4 Vocal Reproduction...81

6.4.1 Key Press versus Sung Direction..81

6.4.2 Sung Frequencies of the Reference Tones...82

6.4.3 Sung Frequencies of the Test Tones...83

6.5 Discussion...83

Chapter 7: Conclusion...93

7.1 Overall Conclusions...96

7.2 Limitations and Future Directions...98

References...101

LIST OF TABLES

Table 4.1: Frequencies (Hz) of the Overtones Used to Imply the Missing Fundamental of the Tones as Function of Spectral Series (Low, Medium, and High) and Semitone (Note on the Piano Scale)..62

LIST OF FIGURES

4.1 Example of a piano keyboard outlining the placement of C4……………………….59
4.2 Schematic of experimental design of the spectral series for Experiment 1…………..61
5.1 Mean proportion correct—concordant with F_0—as a function of Test Condition, Spectral Series, and Piano Key number for both musicians and non-musicians in Experiment 1..73
5.2 Mean proportion correct—concordant with F_0—as a function of Counterbalancing Condition, Spectral Series, and Piano Key number for the initial Test condition of both musicians and non-musicians in Experiment 1…………………………………………..74
5.3 Key-Press versus Sung Direction in the Vocal Reproduction Condition of Experiment 1……………………………………………………………………………………..75
5.4 Sung Frequency of Reference Tone - Tone F_0 (corrected for octave) in the vocal reproduction condition of Experiment 1...76
5.5 Sung Frequency of Test Tone - Tone F_0 (corrected for octave) in the vocal reproduction condition of Experiment 1...77
6.1 Mean proportion correct—concordant with F_0—as a function of Test Condition, Spectral Series, and Piano Key number for both musicians and non-musicians in Experiment 2..88
6.2 Mean Proportion Correct Sung Direction in the Vocal Reproduction Condition for Musicians and Non-Musicians in Experiment 2……………………………………...…89
6.3 Mean Proportion Correct Key-Press versus Sung Direction in the Vocal Reproduction Condition for Musicians and Non-Musicians in Experiment 2..…………………………90
6.4 Mean Sung Frequency (Hz.) of Reference Tone - Tone F_0 (corrected for octave)……………………………………………………………………………………91
6.5 Mean Sung Frequency (Hz.) of Test Tone - Tone F_0 (corrected for octave)………..92

CHAPTER ONE: INTRODUCTION

Research regarding pitch perception has spanned a broad domain of research areas, including anthropology, neuroscience, fine arts, and psychology faculties (Balaguer-Ballester, R. Clark, Coath, Krumbholz, & L. Denham, 2009; Shamma, 2004; Song, Osmanski, Guo, & Wang, 2016; Lui et al., 2016). In particular, research regarding the role of fundamental frequencies and overtones has only recently emerged due to the creation and availability of spectrograms (visual representations of sound via audio waveform) and the ability to extract tones through computer programs such as Praat (Boersma & Weenink, 2011). Praat analyzes the recordings of speech (or singing) and allows the user to parse out specific aspects of sound, such as the fundamental frequency. Of particular interest in the field of pitch perception have been differences in the abilities of musicians and non-musicians in their capacity to identify and reproduce pitches with varying qualities, and to identify what aspects of pitch individuals actually require to perceive pitch.

With the expansion of technology, we are able to manipulate sounds in ways that were previously impossible. The field of researching pitch and overtone perception really began in 1841, with August Seebeck. He created spinning plates with holes drilled into them, so that they would produce tones (or sirens) when they were spun. Seebeck discovered that different combinations of spinning plates would result in hearing the implied fundamental frequency, which will be described in detail throughout this thesis. His polyphonic sirens resulted in the ability to mathematically calculate pre-existing tones and precisely hear the pre-existing series of overtones produced, paving the way for future researchers to test the calculations of tones and perception of overtones (Pinch & Bijsterveld, 2012).

1.1 SOUND

1.1.1 Pitch

Pitch, or sound frequency, is simply a categorical assigned name we give to tones. Gestalt psychologists have outlined a series of grouping rules that explain how we perceptually classify the world around us (including sound) (Cherry, n.d.). The rules state that humans have a tendency to group like elements together through pattern identification and the simplification of complex stimuli (Cherry, n.d.). Based on these grouping principles, the term "pitch" is more a term that more describes the grouping of similar frequencies into categories. For example, the frequency of the note B3 is 246.94. The next note in the sequence, C4, has a frequency of 261.63. When we listen to these notes, we perceptually classify frequencies closer to 246.94 as B3 (though we may note they are sharper or flatter than a "true" B3) and frequencies around 261.63 would be classified as C4 (again, sharper or flatter than a "true" C4).

Tones exist on a continuum, and we perceptually classify specific auditory vibrations into pitches. Faster vibrations result in higher pitches, and slower vibrations result in lower pitches. The classifications of these tones and their associated *pitch height* is universal (Shepard, 1964; Krumhansl, 1990). Pitch height is defined as the "highness" or "lowness" of a pitch in relation to frequency whereas pitch chroma refers to the position of the pitch within a defined octave (Plack, Oxenham, & Fay, 2005).

A common analogy to pitch (and the arbitrariness of its classification) would be the identification of colour. Colour exists on a spectrum, yet we tend to group colours together. Lilac and plum are both descriptives of the colour purple, but where does that distinction occur? When does lilac become plum, and how do we decide that plum is purple instead of blue? As with colour, there is a tendency to classify similar frequencies in like groupings, which Gestalt psychologists have identified as one of the grouping principles and named the *Law of Proximity* (Coren & Girgus, 1980). In regard to audio frequency, sounds that are similar in frequency are more likely to be classified as the same, even though they may differ in terms of their specific wavelengths, just as plum and lilac would both be identified as purple. Therefore, it is arbitrary to label pitches as "specific sounds," because a grouping of a pitch may include a number of different frequencies that we perceive as the same. Giving these sounds a name (specifically, a pitch name) enables individuals to judge sounds as "higher" or "lower" (Plack et al., 2005).

What distinguishes pitch from simply noise is the ability to clearly distinguish a stable enough sound (Powers, 2003). Pitch is directly correlated with fundamental frequency, as the fundamental frequency is the lowest perceived tone of a pitch. Pitch is integral in human communication and music (Hoeschele, 2017). In certain languages, such as Mandarin, pitch fluctuation is crucial in deriving meaning. In all known languages, for example, an increase in pitch at the end of a sentence indicates a question (van Heuven & van Zanten, 2005). Pitch can also oftentimes be attributed to physical characteristics, thereby giving the listener more information as to the direction or originator of the sound (Hoeschele, 2017). For example, the F_0 of speech can provide information regarding the gender of the speaker; specifically, it is a necessary element needed to identify the relative pitch level as well as the gender of the speaker (Bishop & Keating, 2012).

1.1.2 Auditory Scene Analysis

Pitch is one of many perceptual attributes of sound that allows us to differentiate between different noises in the environment. For example, pitch allows us to differentiate between sound sources in the same environment, such as two people speaking at once (Hall & Palack, 2009). We are constantly being bombarded with auditory information. The auditory system needs to be able to segregate components of sound from different sources and group auditory information from the same sound source together. The field of auditory scene analysis aims to understand how the auditory system manages and distinguishes the perceptual overload of auditory stimuli from different sources and the ability to organize sound in perceptually consequential components (Bregman, 1990).

Bregman (1990) integrated two sequential ways in which we analyze auditory stimulation from different sources. Multiple sounds from several sources can either be “integrated” into a whole sound (for example, a full-sized orchestra playing a piece of music) or “segregated” into individual components (such as a conductor only listening to a clarinet line in a full orchestra). Linking segregated sounds together produces an auditory stream, allowing the conductor to be able to identify the auditory stream of the clarinet line as distinct from the rest of the orchestra. The linking of segregated sounds is known as the *cocktail party effect*, as a common example is being able to attend to one speaker while several other conversations are occurring around you at the same time (such as at a cocktail party) (Bregman, 1990).

The Gestalt grouping principles come into play in regard to Auditory Scene Analysis. In an evolutionary sense, sounds that had similar frequencies were likely to come from the same, if not similar, sources. For example, early hominids did not need to be able to distinguish the individual roars of a lion; all they had to know was that a roar of a lion meant a lion (or a pride) was nearby. This applies to the second way Bregman (1990) proposes in the perception of sound; segregation. Segregation is defined as our ability to recognize patterns of perceptual cues.

Forming auditory streams has a significant impact on our ability to perceive auditory stimuli. Streaming is almost a form of “competitive grouping”; more similar streams or sounds are more likely to be grouped together and dissimilar groups will diverge into their own streams (such as the brass section versus the woodwind section of an orchestra). Audience members of an orchestra performance are more likely to experience a “big picture effect” caused by competitive grouping, as opposed to following a single musical line (such as the trumpet line). Of course, composers may emphasize certain instruments throughout a piece, but overall the effects of the orchestra playing simultaneously is far easier to perceive than 50 individual instruments at once.

Music is defined as the sequence and organization of specifically chosen pitches, along with other musical attributes such as timbre, duration, and loudness (Oxenham, 2012). However, it is important to note the differences between pitch and timbre, as they are often assimilated and thought of as one spectral property of sound. Pitch is associated with the fundamental frequency and associated overtones (more of which will be described further into this thesis). Timbre, on the other hand, is a much more complex auditory attribute that has been described as a perceptual fusion of auditory events (Siedenburg & McAdams, 2017). Timbre has more to do with the identification and specification of a sound source than the actual tone being played, as well as giving sound its quality (Hoeschele, 2017). For example, a clarinet and a piano can play the same phrase in a piece of music and their pitches will be identical; however, their timbre is drastically different. Pitch has more to do with the frequencies produced by the actual tone and less to do with the source of the tone.

1.1.3 Frequency

Frequencies are measured sound waves that occur when a sound is articulated. A frequency is defined by how many waves pass in a given place or time, indicating the speed of vibration and the pitch. Amplitude determines how loud the sound is by the size of the vibration. (*Sound and music*, 1997-2019). Frequency is measured in Hertz (Hz). A frequency of 10 Hz is equivalent to ten wave cycles per second. The longer the wave cycles, the lower the perceived pitch (and frequency) will be.

1.1.4 Fundamental Frequency

Fundamental frequencies, often called fundamentals or F_0, are the lowest perceived note or tone of a musical or linguistic pitch. Quite simply, they are the tones heard when someone is speaking or an instrument is being played. A single fundamental produces a harmonic series, also known as the overtone series. Since fundamentals are the lowest perceived note, they are the first tones in a harmonic series. Subsequent notes are overtones, and the fundamental would not be included in that grouping. In a mathematical sense, overtone frequencies are integer multiples of the fundamental. They are the associated frequencies above the note heard, and occur for musical instruments as well as speech (Classroom, 2018).

For example, the frequency of C3 (the fundamental frequency in this example) is 130.81 Hz. Following this, the next tone in the harmonic series is C4, with a frequency of 261.63 Hz. This note is perceptually one octave higher than C3 (octaves are defined by the ratio 2:1; the frequencies (in Hz) of the subsequent octaves will always be double their predecessor). The next tone in the series is G4, with a frequency of 391.99 Hz and a fifth above C4 (a perfect fifth has a frequency ratio of 3:2). Regardless of what the fundamental frequency is, the overtones will always be harmonically related as they are integer multiples of the fundamental.

Pure tones do not exist in nature. A completely pure tone is a Hertz frequency (Hz) that forms an S-shape sine wave and is audibly a single tone (lacking the associated harmonics correlated with complex tones). Frequency (Hz), amplitude, power, and phase (time in relation to other sound waves) are all elements included in the mathematical formation of a sine wave (*Pure Tones*, n.d.). Oftentimes, pure sine waves can be used as a foundation for more complex tones. Pure tones cannot be produced vocally or by any instruments, but they can be generated electronically and help us understand how the perception of auditory stimuli occurs. Previous research has indicated that it is more difficult to localize pure tones than tones containing more auditory information (Stevens & Newman, 1936).

Fundamental frequencies are included in the makeup of complex tones, which are combinations of two or more pure tones (Chialvo, 2003). They are responsible for the "pitch" heard when listening to complex tones. The associated tones are harmonics which are integer multiples of overtones of each fundamental frequency (Hoeschele, 2017).

1.1.5 Harmonic Analysis

This branch of mathematics includes the study and understanding of the *Fourier Series* and the *Fourier transforms*, both of which are branches of the *Fourier Analysis*.

The *Fourier Analysis* is described as the process of breaking down complex tones into the F_0 and associated harmonics in order to identify the harmonic content of a sound, such as the quality or timbre of the original sound. This branch of mathematics in assessing pitch perception was discovered in 1822, when Joseph Fourier demonstrated how a potentially infinite sum of harmonics could be deconstructed into certain functions (Fourier, 1822).

Fourier Series are a way to represent a function of sound as the sum of simple sine waves using *periodic* and *continuous* time domain signals (S. W. Smith, 1999). Continuous waveforms are infinitely repetitive harmonics (also known as overtones) and periodic signals are domains that include harmonics, with a value of zero for the frequencies between harmonics (as they do not contribute to the time domain signal). The *Fourier Transform* is described as: "transforming a time-domain signal to the frequency domain" (Truax, 1999). This transform can be computed electronically, but it seems the human ear is also quite capable of separating and identifying harmonics (see section titled *Pitch Perception* for more details). In other words, the *Fourier Transform* is the relationship between time and frequency.

The *Fourier Analysis* combines the fundamental frequency and associated overtones, which are then known as partials. The combination of simple sine waves allows for the formation of more complex and harmonic sounds, as they are integer multiples of a certain frequency. The discoveries of these types of analyses have fundamentally changed the way pitch perception research has been conducted. These analyses allow researchers to manipulate tones and to control what they want participants to hear. As a result, it is much easier to research specific aspects of sound as well as our reactions to them.

1.1.6 Overtone Series

Overtones can be subtly heard, and they occur for (almost all) naturally produced notes or tones. Harmonics are integer multiples of overtones of each fundamental frequency (Hoeschele, 2017). For example, when a cymbal is struck, both the fundamental and the overtones can be heard. Another example in which the harmonic series and fundamental can be heard is in polyphonic singing (commonly known as throat singing). This type of singing, found in many cultures around the world, is a popular method in which the singer manipulates the resonances produced to create more than one perceived pitch at the same time (Creighton, 2014). Peoples of Tibet and Siberia, known as the Tuva people, make use of polyphonic singing. One singer seems to be producing two pitches at the same time (giving way to the appearance of singing two notes) when in reality they are producing a single F_0 that acts as a background drone; the second "note" is made up of selected harmonics (of the projected F_0) that are accentuated (Plack et al., 2005).

1.1.7 The Phenomenon of the Missing Fundamental

The phenomenon of the missing fundamental refers to the perception of a complex pitch constructed without the F_0, while possessing the ability to hear the "missing" pitch (Plack et al., 2005; *Missing Fundamental Effect*, n.d.). This phenomenon is universal across all cultures and has been well-documented in the animal kingdom (Qin, Sakai, Chimoto, & Sato, 2005). If the lower harmonics and F_0 are not produced, we are still able to "hear" the correct tone. Beat frequencies are the speed at which different waves of the associated harmonics vibrate, and the beat frequencies between the associated harmonics help to reinforce the pitch of the F_0, even if it is missing (*Missing Fundamental Effect*, n.d.). The beat frequency of the associated harmonics corresponds to the missing F_0. Although there is no true source of vibration of the F_0, it can still be “heard” and perceived due to beat harmonics.

Previous research has shown that certain regions of the primary auditory cortex of the auditory system (specifically, the temporal areas of the neocortex) play a role in the missing fundamental (Zatorre, 1988; Matsuwaki et al., 2004). In particular, Heschl’s gyrus and surround- ing cortex in the right cerebral hemisphere play crucial roles in identifying and calculating the missing fundamental tone from a complex tone.

When the F_0 is missing, other factors such as frequency, intensity, duration, and harmonics need to be taken into account in order to perceive the missing pitch. Frequency is defined by the number of waves passing in a given place or time and indicates the pitch (*Sound and music*, 1997-2019). Duration refers to the amount of time a pitch needs to be expressed in order to properly perceive a clear sense of the tone (Plack et al., 2005). Depending on the frequency and intensity, a pitch needs to last for around 10-60ms (Plack et al., 2005). Sound intensity, measured (for hearing) in decibels (dB), is defined as the sound capability, often based on the listener’s location (*Sound Intensity*, n.d.). Frequency, intensity, duration, and harmonics can all give clues as to what the missing pitch would be.

1.2 PITCH PERCEPTION

Humans have certain neurological processes that are able to differentiate pitch from other aspects of sounds, such as timbre. On average, humans can hear a range from about 20 Hz to 20,000 Hz (Cutnell & Johnson, 1998). Humans as well as animals are able to process and assess the sound waves of pitch through pitch height, pitch chroma, relative pitch, and grouping principals, although auditory perception can be context-dependent (Sussman & Steinschneider, 2006).

Pitch height refers to the "highness" or "lowness" of a pitch in relation to frequency whereas pitch chroma refers to the position of the pitch within a defined octave (Plack et al., 2005). Relative pitch is defined as the ability to identify the direction or interval (or sequential frequency ratio) of a pitch change when given a reference tone (Hoeschele, 2017). Relative pitch is relatively common among musicians, especially those who are able to "play by ear." Truax (1978) noted perceptual differences of pitches on the basis of the spectrum of overtones.

An abundance of overtones strengthens the perception of the fundamental; in contrast, the more "pure" a tone sounds (and the more the waveform resembles an S-shape sine wave), the more distinct the F_0 is, yet the ability to accurately perceive the pitch begins to deteriorate. It seems that auditory perception relies (to some extent) on the relationship between the F_0 and overtones. However, humans appear to have a weak sense of pitch height. It was even noted that possessors of absolute pitch had no better judgments of pitch height than those lacking absolute pitch (Takeuchi & Hulse, 1993; Deutsch & Hawthorne, 2004; Lockhead & Byrd, 1981). Instead, those with absolute pitch alternately rely on pitch chroma. In fact, it has been observed that most humans are able to attend to pitch chroma, as can be seen by the reproduction and comparison of familiar melodies (Levitin, 1994; N. Smith & Schmuckler, 2008; Schellenberg & Trehub, 2003).

Previous research has shown that both mammalian and avian species are able to perceive pitch, even with stimuli constructed without the F_0 (Cyx & Shapiro, 1986; Heffner & Whitfield, 1998). There has also been research suggesting an ability to perceive tones missing the F_0 in human infants (aged 3 to 4 months) (Lau & Werner, 2012). In an evolutionary sense, even being able to perceive parts of pitch allows animals (including us) to identify key information. For example, when a male is speaking on the phone and poor reception causes interference or feedback, the listener is still usually able to identify the speaker as male (by his lower pitch), even though some of the spectral information (such as accent, word usage, speed, and spacing) is lost due to the poor reception.

Certain aspects of pitch begin to be perceived as sound waves reach the ear. The outer ear (pinna) is shaped in such a way that funnels sound into the external auditory canal. As the sound waves travel through the external auditory canal, they cause a series of tiny bones in the middle ear to vibrate, turning the sound waves into mechanical energy. The mechanical vibrations of these bones, known as the malleus, incus, and stapes, collide with the fluid-filled cochlea, converting mechanical energy into hydraulic energy (Centres, n.d.). There are approximately 15,000 hair cells along the membrane of the cochlea; their main purpose being transmitting the hydraulic energy of the endolymph (cochlear fluid) into a neurological signal that is able to be processed by the brain (Pujol, Nouvian, & Lenoir, n.d.). Hair cells at the base of the cochlea (closest to the bones) are responsible for higher frequency pitches, and the hair cells furthest way from the bones are responsible for lower pitches (Centres, n.d.). The frequency-place code of the cochlea (specifically, the basilar membrane) results in a spatial arrangement of tones in the auditory cortex known as a tonotopic map (Saenz & Langers, 2014). We are able to perceive pitch once it is expressed along this map, and this phenomenon is known as the place theory of pitch.

Another theory of pitch perception states that the perception of sound depends on temporal patterns of neurons in the cochlea responding to sound. The temporal theory determines the pitch of a pure tone through the transient firing of neurons (either single neurons or as a group of neurons, also known as the volley theory). As opposed to the place theory, the temporal theory states that regardless of the place of the firing rate, the timing of neuronal firing is indicative of pitch perception (Plack et al., 2005).

A sense for music seems to be a universal human trait found among all cultures (Hoeschele, 2017). In particular, the use of pitch (particularly, discrete pitches and simple pitch intervals) is one of the most common universals found in human music, from pitched drums, to singing, to other forms of instruments. In language, pitch helps to convey meaning and can convey semantic and emotional information (Hoeschele, 2017). Prosody is defined as perceptual elements of speech that are not individual phonetic articulations or utterances (Nooteboom, 1997). These perceptual elements include voice pitch, intentional pitch fluctuation, and the length and loudness of the oration.

1.2.1 Stimuli in Pitch Perception Research

The tones used in pitch perception tasks often have specific features created electronically, as previously mentioned in the "Harmonic analyses" subsection. When tones are presented to participants, certain auditory features may be included or excluded. For example, overtones may be manipulated or the F_0 may be removed in order to assess what aspects of pitch are necessary for perception. Often, the onset and offset of tones will be manipulated as an experimental control, allowing participants to receive the stimuli in a controlled manner.

1.2.2 Individual Differences

There are differences in individual abilities to perceive pitch. For example, specific early childhood musical training encourages the development of absolute pitch (commonly known as perfect pitch), in which musicians are able to accurately identify the names of pitches (such as a B flat) due to superior *pitch memory* without hearing a reference pitch (Levitin, 1994; Hoeschele, 2017). An important aspect of absolute pitch is that individuals do not remember the actual tone, but have associated a verbal pitch name to the sound (Takeuchi & Hulse, 1993). The ability to possess stable auditory imagery of pitch representations is a remarkable feat that would aid an individual in pitch discrimination tasks. Not all musicians develop absolute pitch, but over time, it seems consistent exposure to music (with an emphasis on interval differences) enhances their ability to distinguish pitch since they are more easily able to identify musical sounds, as opposed to individuals who do not have constant exposure and specialized training (Hamilton, 2009; Welch, 1979). It is said that this rare ability occurs in one in 10,000 people (Bachem, 1955).

Although true absolute pitch is a rare phenomenon, those without this ability are still able to discriminate pitch. For example, Latinus and Taylor (2012) noted that absolute pitch was the first piece of information used when discriminating the gender of speakers. Levitin (1994) also showed that individuals still possess some form of absolute pitch. Participants were asked to sing a popular song from memory and about two thirds of the participants were able to sing within the accuracy of two semitones with no prompting. In other words, they were fairly close to singing the popular song with 100% accuracy. N. Smith and Schmuckler (2008) tested individuals who did not possess absolute pitch and found that they were able to accurately identify a commonly heard tone, such as a dial tone. Although we all seem to possess some form of absolute pitch, true abilities lie with those who are able to not only identify tones but also accurately label them.

A much more common phenomenon found among humans is the notion of relative pitch; the ability to identify the direction or interval of a pitch change when given a reference tone (Hoeschele, 2017). For example, an individual with relative pitch would easily be able to identify the direction of an A note followed by an E note. Relative pitch allows us to identify common songs, such as “Happy Birthday,” regardless of what note the tune begins on. It is relatively easy to train individuals to identify pitch direction, such as teaching an individual what the interval of a “perfect fourth” sounds like and how to identify it. Typically, an individual will be instructed to associate the introduction of certain melodies with intervals; for example, Amazing Grace begins with a perfect fourth. In certain languages (such as English and German), relative pitch plays a key role in identifying stress between words or phrases (Kohler, 2012). In more tonal languages (such as Mandarin and Vietnamese), pitch fluctuations within a word (as opposed to between words) can change the meaning (Ngo, Vu, & Strybel, 2016).

Another extreme difference in individual ability to perceive pitch is the notion of tone deafness, or amusia. Amusia, by true definition, means lack of music. Individuals with amusia have a marked inability in identifying or reproducing pitches. Whereas individuals who possess absolute pitch have a keen memory for pitch names, amusics have an impaired memory for pitch due to congenital effects or brain damage later in life (Peretz et al., 2002). Although it may seem that tone deafness is common, it is reported that only two to five percent of the population possesses true amusia (O’Connor, 2019).

Differences in brain structure result in individual differences in sound perception. An excess of cells in the hearing centre of the right cerebral cortex (also referred to as

Heschl's gyrus) generate more clearly heard overtones, allowing an individual to process the overtones more clearly. However, individuals with an excess of neuronal cells in the hearing centre of the left cerebral cortex are able to hear the fundamental more clearly (Schneider et al., 2005). This study resulted in the emergence of listening types; both "overtone listeners" and "fundamental listeners," solely on the basis of participant preference and brain structures measured with magnetoencephalography (MEG) (Schneider et al., 2005). MEG is defined as the measurement of magnetic fields brought about by the electric currents in the brain, providing an accurate and non-invasive portrayal of activity (Singh, 2014). Being attentive to individual differences (especially between musicians and novices or non-musicians) will help in our understanding of how pitch is perceived (Youngdahl, 2018).

1.2.3 Pitch Reproduction

Pfordresher and Brown (2009) have demonstrated that an individual's native language influences their ability to imitate or discriminate musical pitch. Specifically, tonal language speakers show advantages in pitch processing tasks that are not related to language. Tone languages, as defined by Pfordresher and Brown (2009), are languages in which word meaning is conveyed via pitch height and/or pitch changes. A common tonal language often cited is Mandarin, as the syllable /ma/ can mean different things based on the tone used. In contrast, intonation languages do not use pitch height and changes to convey meaning. For example, Malay (the main language spoken in Malaysia and Indonesia) is classified as a non-tonal language. In Malay, syllables are strung together to form sentences, and all syllables are pronounced equally (Zainal, Hussain, & Samad, 2012).

Vomberg (2017) attempted to track where the possible difference between musicians and non-musicians lay in a series of pitch discrimination tasks with tones constructed without the F_0. In this study, participants were placed in one of two contexts; music or speech. In the music context, participants were asked to identify whether the test tone went up or down. All of the instructions for the music context consisted of musical terms such as “pitch,” “tones,” “up,” and “down.” In the speech context, all of the instructions of the study used common speech terms. That is, instead of using words such as “up,” “down,” “pitch,” or “tone,” participants were instead asked to assess the articulation of what they heard based on prosodic cues. They were told the sounds they were about to hear were recordings of someone speaking a two-syllable word with the meaning digitally removed by computer software. Their task was to judge how confident or questioning the speaker’s pronunciation of the words were, and instead of pressing keys for “up” or “down,” they would indicate their response by pressing “questioning” or “confident.” Both contexts had two different conditions. In the *articulate* condition, participants were asked to either sing or hum the two tones (in the music context), or articulate a two-syllable word (in the speech context) before pressing the computer keyboard to indicate their judgment. In the *no-articulate* condition, participants were simply asked to indicate their response by computer key-presses.

In the music context, the accuracy of the vocal direction was almost perfect for musicians and non-musicians. However, musicians were more accurately able to translate their vocal performance into a perceptual judgment (key-press response). Non-musicians in the music context demonstrated a marked dissociation between their key-press perceptual judgements and their vocal pitch directions - the latter of which were generally more accurate, and both musicians and non-musicians demonstrated the dissociation to some extent in the speech context. This dissociation is consistent with the hypothesis that at least one possibility for the difference between musicians and non-musicians in the music context is not that the perceptual experience of non-musicians is different from that of musicians, but rather that non-musicians had difficulty translating their perceptual experience of an implied fundamental frequency into the correct perceptual judgement - a task that the musicians were able to easily perform. It seems that presenting pitches constructed without the fundamental to non-musicians seems to have an effect on their ability to correctly identify the direction of tone pairs, although this does not affect musicians. It is of great interest to this thesis to attempt to find out what musicians are doing when they make a translation from two notes (the tone pairs) to a perceptual judgement, and attempt to teach it to non-musicians.

There are a few theories pertaining to why musicians are better able to complete this task. A common idea is that musicians are simply different. That is, they are born with special auditory systems that draw them to music, as they find it compelling and interesting and ultimately want to participate in musical activities. They fundamentally hear the world differently, and this difference is innate and built in. Makin (2014) stated that these differences were due to genetics; that is, musicians genetically were better at detecting pitch differences and durations. The genetic variation found in musicians showed more acute development of the inner ear and inferior colliculus, and another specific gene that played a role in the ability of hair cells to convert sound into electrical signals. Schneider et al. (2002) used MEG measurements to demonstrate marked brain mass differences between musicians and non-musicians, particularly in the primary hearing centre. They have also shown a stronger reaction to sound in the brains of musicians when compared to non-musicians.

Another possibility is that musicians are born with the same ears as non-musicians, but the musical training throughout their life changes the way musicians hear and perceive the auditory world. It has been shown that musical training improves hearing and the ability to remember sounds (Hamilton, 2009). Zatorre (2003) also demonstrated that specific early childhood training (during the critical period, from around age 5 to puberty) can result in absolute pitch, which is the ability of musicians accurately identifying the names of pitches due to superior *pitch memory* without hearing a reference pitch (Levitin, 1994).

Musical training seems to be context-dependent. That is, training to think in musical terms as opposed to linguistic, even though both domains use pitch to convey different types of meaning, is dependent on context. Non-musicians should still be comfortable with pitch fluctuations, since most languages make use of prosody to convey meaning. Prosody conveys meaning through rhythmic and pitch fluctuations, and although we are not explicitly aware of it, most tone language speakers have the ability to detect pitch (Pfordresher & Brown, 2009). In tone languages, variations of pitches distinguish words and help to convey meaning (*Tone Language*, n.d.). For example, prosodic fluctuations and cues can be indicative of a question, independent of grammar or sentence structure. The use of intonation, pitch height, and pitch contour help to convey meaning (Selting, 1992). An exaggerated example of using pitch fluctuation to convey meaning can be noted by listening to “Valleyspeak.” This phenomenon occurred in California during the 1980’s and 1990’s and was defined by rising intonation (also known as “uptalk”) (Woo, 2013). However, it should be noted that “Valleyspeak” does not specifically convey meaning; it is instead an example of exaggerated prosody and speaking more to group membership.

If non-musicians were unable to perceive pitch changes, they would be unable to communicate effectively. Regardless of whether musicians are innately different or not, studies looking at pitch perception with or without the fundamental frequencies all have similar conclusions; auditory perception seems to be flexible, and not an innate built-in ability. What this thesis aims to do is find out what musicians are doing that helps them complete pitch perception tasks and why non-musicians struggle, and to discover ways in which we can improve the performance of non-musicians in a pitch discrimination task.

1.2.4 Tonal Memory

Tonal memory is an important aspect to consider when asking participants to reproduce pitches unaided. Tonal memory is defined as the ability to recall the tone of a previously sounded pitch without external aid (Gorow, 2011). Musicians with absolute pitch are able to perfectly identify and reproduce pitches heard without a reference pitch (Levitin, 1994). However, they may not have an exceptional memory for the tone itself, but simply be able to remember the name assigned to the pitch they heard.

1.2.5 Vocal Range

Acknowledging individual differences in vocal ranges is important when conducting research requiring participants to reproduce pitches. Vocal range is the range of tones (the perceived ”highness” or ”lowness”) an individual can produce vocally. Longer vocal cords produce lower pitches, as they vibrate more slowly. As shorter vocal cords are thinner and vibrate faster, they produce higher pitches (Klofstad, Nowicki, & Anderson, 2012). Hormones and larynx size help determine the size of vocal cords. When speaking, the pitch of the voice is the fundamental frequency, which is established by the flow of air causing the vocal cords to vibrate.

Vocal range can be somewhat malleable (with sickness, age, and environmental factors such as pollution), although a majority of it is determined by genetics (specifically, the formation of vocal folds) and hormones. Often underestimated, the average vocal range is 3 1/3 octaves, encompassing about forty notes (Popeil, 2015). For males, the average vocal range occurs between 85 hertz to 180 hertz; for females, the average pitch range is from 165 hertz to 255 hertz. Testosterone plays a major role in the different ranges; in particular, testosterone has a major effect on the male larynx (commonly referred to as the “Adam’s apple”), resulting in a lower voice pitch (Klofstad et al., 2012). It is important to note that this range and the malleability mentioned are found in healthy individuals with no previous voice disorders. Many individual factors, such as substance use or traumatic brain injuries, can affect vocal ranges.

When talking about vocal range, it is important to consider vocal registers; the chest voice and the head voice. The chest voice is analogous to the speaking voice and is the fullest and most resonant register for a majority of the population. It is also often the “bottom” of the singing range (excluding trained sopranos and tenors). When individuals who are not trained vocalists sing, they often feel most comfortable using their chest voice. In contrast, the head voice is used when singing higher pitches (though not to be confused with falsetto). It has been argued that there is also a third register, bridging the head and chest voice. This “mixed” voice is where most individuals feel their voice cracking or a vocal strain (*Singing 101: Ranges and Registers*, 2017). When singing along to the radio, for example, it is most likely that the singer (if they are untrained) will transpose the music to fit within their chest voice, as that is the range they are likely most comfortable in (Grody, n.d.). However, there are uncommon but important individual differences that need to be taken into account when discussing vocal ranges as well as pitch reproduction or discrimination tasks. If participants are asked to reproduce tones, the tones will need to fall in the range most comfortable for the majority of participants (likely the chest voice).

1.3 INFLUENCE OF LANGUAGE

When we speak (any language), or play an instrument, we are exposed to the fundamental and the concurrent harmonic series. How then, for instance, are we able to relay information and understand responses, and ultimately perceive the pitches being presented to us? An evolutionary theory that attempts to explain this is the Music-Language hypothesis. This hypothesis describes pitch perception and production, and how we have evolved to pay attention to certain cues. This hypothesis states that modern-day music and language evolved from a common ancestor of communication known as the "musilanguage stage" (Brown, 2000). The theory behind this stage is that the structural features shared by music and language emerged simultaneously and evolved to have different perceptual attributes. In music, emotive meaning (also known as social meaning) is emphasized and expressing emotions or causing strong feelings is a key feature. In contrast, referential meaning is emphasized in language, defined by the analytical, cognitive, and indicative content (Brown, 2000; Ogden & Richards, 1923). The use of pitch helps to convey these different types of meaning, and how individuals perceive pitch helps them to respond appropriately. We have evolved to hear certain pitches in certain ways, and our cultural development helps shape who we are as communicators and musicians.

There is an important parallel that can be made here between infant language and musical development. Language development is categorical. Infants can hear and distinguish all the language sounds in every language, and they tune their perception to match characteristics in their surroundings. As children develop (within the first few years of life), they learn to distinguish across categories and generalize within a category, implicitly learning what is and is not important (Werker & Yeung, 2005). An example of categorical organization in language would be phoneme differences. Phonemes, in any language, are sets of speech sounds that distinguish one word from another, and are the building blocks of language (Collins, 2018). For example, in the English language, the letters *P* and *B* are used to convey different meanings (think *pat* or *bat*). In contrast, phonemes do not play as crucial a role in Mandarin. Instead, syllables and pitch height (or variations on phonemes) are used to convey meaning (Ly, 2012). Understanding the boundaries of these categories is crucial for language development and understanding.

In theory, speech sounds are not categorical. Sound, including speech sounds, occur on a continuum. For example, at some point the sound *ba* will morph into the sound *pa*. However, it has been demonstrated that when people are presented with these sounds on a continuum, they will place the sounds into different categories (Liberman, Harris, Hoffman, & Griffith, 1957; Harnard, 2003). Although Liberman et al. cited this "categorical perception" as a speech-specific phenomenon, it can be argued that the same phenomenon occurs in music. Western music, at an everyday level, is categorical. Pitches are classified as specific notes, which are then categorically placed in a repeating pattern (known as scales). However, when analyzing notes in a spectrogram, pitch is continuous. Perhaps, like language, generalizations are made within musical categories, and when presented with ambiguous pitches (from a spectrogram), they will always be assigned to a category.

1.4 PREVIOUS RESEARCH

Previous research has shown that non-musicians could accurately identify small pitch changes when the fundamental frequency was present in the pitches (Tervaniemi, Just, Koelsch, Widmann, & Schroger, 2004). The next step in this research was to present pitches constructed without the F_0 to participants in a series of tone pairs. Tone pairs consist of two different tones that are presented to participants. The first tone is often considered the "reference" tone, which is then followed by a "test" tone. The two tones will often fall within a range (for example, two tones within the span of an octave). In a pitch discrimination task, participants are asked to differentiate between the tones. Discrimination tasks are the most common way to conduct a pitch perception task. Participant are asked to choose whether the test tone is higher or lower than the reference tone. Oftentimes, participants are required to reproduce (usually by singing or humming) the tones presented to them (Welch, 1979; Tervaniemi et al., 2004). As previously stated, overtones are simply integer multiples of the fundamental. Therefore, the pitch of a tone will remain "somewhat" constant, even after the fundamental is removed (Oxenham, 2012). It has been of interest to researchers how constant and easily perceivable the remaining overtones are to participants (both musicians and non-musicians).

Seither-Preisler, Krumbholz, Patterson, Johnson, and Nobbe (2007) were among the first to conduct a pitch perception task measuring the differences between musicians and non-musicians using tones constructed without the fundamental frequency. The following chapter analyzes their study and results, followed by the influence they had on further research.

The tones used in pitch perception tasks often have specific features created electronically, as noted in the “Harmonic analyses” subsection. When tones are presented to participants, certain auditory features may be included or excluded. For example, overtones may be manipulated or the F_0 may be removed in order to assess what aspects of pitch are necessary for perception. Often, the onset and offset of tones will be manipulated as an experimental control, allowing participants to receive the stimuli in a controlled manner.

CHAPTER TWO: THE PHENOMENON OF THE MISSING FUNDAMENTAL: SEITHER-PREISLER ET AL. (2007)

As mentioned in the previous chapter, there has been research conducted regarding the phenomenon of the missing fundamental frequency. Seither-Preisler et al. (2007) are some of the first to study this phenomenon in a pitch perception task, comparing the performances of musicians and non-musicians. Seither-Preisler et al. (2007) used the Auditory Ambiguity Test (AAT) to identify differences in pitch perception with tones constructed without the F_0. Participants were classified into three groups: professionals, amateurs, and non-musicians. The groups were presented with tone pairs that were constructed without the F_0. Based on performance accuracy, the groups were then dichotomized into “spectral” listeners or listeners who attended to the missing F_0. “Spectral” listeners were able to perceive the tones as a whole, even though the tone were constructed without the fundamental frequency. “Spectral” listeners tracked the overtone series as opposed to the implied F_0 (as tracked by the missing-F_0 classifiers).

2.1 PARTICIPANTS

2.1.1 Professionals, Non-Musicians, and Musical Amateurs

Seither-Preisler et al. (2007) defined professionals as individuals who had received formal music education through a music conservatory and who engaged in regular practice. In total, 18 professionals participated in the study (*M*=31.2 years of age, *M*=23.8 years of musical practice; 11 women, 7 men). Non-musicians were described as participants who have not played a musical instrument past the age of 10 years. There were 30 non-musicians tested (*M*=30.9 years of age; 23 women, 7 men). Musical amateurs were identified as having a limited musical education (at minimum, one hour a week) and who had practiced more than one instrument in the past year. In total, 31 amateurs were tested (*M*=28.6 years of age, *M*=12 years of musical practice; 24 women, 7 men).

Seither-Preisler et al. (2007) acknowledge the imbalance among participant groups, opting to reduce the number of amateurs and non-musicians. The authors only included participants with low guessing probability. This will be discussed in greater detail near the end of this chapter.

2.2 AUDITORY AMBIGUITY TEST (AAT)

The AAT consisted of 50 different tone pairs (100 tone sequences) containing frequency shifts with a range of 2 to 9 semitones for the implied F_0. All of the stimuli moved in opposition; if the overtones were rising in pitch, the associated (implied) F_0 would fall in pitch. As well, if the overtones were falling in pitch, the implied missing F_0 would rise in pitch. Each of the tones followed a Gaussian distribution in which the tones ascended by 10 ms, plateaued for 480 ms, and descended with a ramp of 10 ms. The tones were presented in pairs with 500 ms difference between the two tones in the pair. Tone pairs were separated by an interval of 4,000 ms. The AAT was taken twice by participants, resulting in 200 tone sequences.

Tone pairs were randomized and presented to participants in 10 blocks, each containing 10 trials. Each trial of the AAT comprised of two tones with different spectral characteristics. As a result, Seither-Preisler et al. (2007) could not conclude that the effects may have been due to the parameters of the tone sequences used. Regardless of the formation of the stimuli, participants were simply asked to identify the direction in a forced-choice paradigm; that is, if the pitch went up or down. No other information regarding the tones was given to the participants. It was of interest to the researchers to identify if the participants tracked the implied F_0 or the overtones, and to what degree. Scores ranged from 0 (*100 spectrally-based responses*) to 100 (*100 F_0-based responses*).

2.2.1 Spectral Construction of the Tones

The tones used in this study were spectrally constructed in three ways: "(a) low-spectrum tone: 2nd-4th harmonic, high spectrum tone: 5th-10th harmonic, *N*=17 tone pairs; (b) low-spectrum tone: 3rd-6th harmonic, high-spectrum tone: 7th-14th harmonic, *N*=17 tone pairs; and (c) low-spectrum tone: 4th-8th harmonic, high-spectrum tone: 9th-18th harmonic, *N*=16 tone pairs." The tone pairs all followed a similar pattern; the frequency ratio between the high- and low-spectrum tone was always corresponding to an octave (a ratio of 1:2). Tones were normalized "so that they had the same root-mean-square amplitude value" (Seither-Preisler et al., 2007).

Seither-Preisler et al. (2007) considered five frequency separations of the missing F_0s: "(a) ±204 cents (musical interval of a major second or two semitones); (b) ±386 cents (musical interval of a major third or four semitones); (c) ±498 cents (musical interval of a perfect fourth or five semitones); (d) ±702 cents (musical interval of a fifth or seven semitones); and (e) ±884 cents (musical interval of major sixth or nine semitones)." The implied F_0 and associated overtones always moved in opposition; if the implied F_0 ascended in pitch, the overtones would descend (and vice versa).

2.2.2 Procedure

Participants were placed at a computer and were informed (via instructions on the computer screen) that they were about to hear 50 tone pairs (100 tone sequences in total) presented in within-order pairs. Participants were asked to correctly identify the direction of the pitch sequences; whether the sequences were rising or falling. The spectral information of the stimuli was not disclosed to the participants and they received no feedback on the accuracy of their performances. Participants were encouraged to go with their “gut-instinct” and they were allowed to imagine humming or singing the sequences. Participants completed the first block of sequences twice in order to become familiar with the AAT, although only the second presentation was included in subsequent analyses. They then completed the 100 tone sequences twice, as the 50 tone pairs were being assessed four times. The possible scores that could be achieved on the AAT ranged from 0 (*100 spectrally based responses*) to 100 (*100* F_0*-based responses*).

2.3 DATA ANALYSIS

The average AAT scores were not normally distributed, resulting in the use of nonparametric tests. Specifically, the Friedman test, Mann-Whitney *U* test, Kruskal-Wallis test, Spearman rank correlation, and Wilcoxon signed-rank test were used. The AAT scores were derived from the “proportion of trials categorized in terms of the missing F_0s” (Seither-Preisler et al., 2007).

2.4 RESULTS

2.4.1 Effect of Musical Competence

Mean AAT scores differed for each group; 45.9% for non-musicians, 61.6% for amateurs, and 81.6% for professional musicians. The differences between the groups were determined as highly significant. Musical competence played a significant role in whether or not participants were classified as “spectral” listeners or “missing-F_0 classifiers.” Specifically, an increase in musical competency paralleled an increase in the proportion of missing-F_0 classifiers. Non- musicians were consistently identified as “spectral” listeners and professional musicians were the most consistent as “missing-F_0 classifiers.” Amateurs were divided between “spectral” listeners and “missing-F_0 classifiers.” In other words, professionals, for the most part, were able to track the implied F_0.

2.4.2 Effect of Interval Width

Increased interval width played a key role in the ability to make F_0-based judgments. The closer an interval was (say, a major second), the more difficult it was to identify the direction of the sequence of pitches constructed without the F_0. This effect (of identifying the direction of tone pairs constructed without the F_0) was found regardless of musical competence. Mean proportions (collapsed across expertise level) of the intervals were as followed; 43% for a major second, 55% for a major third, 60% for a perfect fourth, 68% for a fifth, and 75% for a major sixth. The accuracy in identifying the direction of the sequences improved dramatically as the intervals increased. This difficulty in identifying small interval differences, especially with pitches constructed without the F_0, has been documented in similar research studies (Granzow, 2010).

Seither-Preisler et al. (2007) again dichotomized participants as “spectral” listeners or “missing-F_0 classifiers” based on their accuracy in regard to each interval in the sequence. Non-musicians and amateurs were more likely to be classified as “missing-F_0 classifiers” compared to professionals, especially as the interval width increased. Specifically, at the interval of a fifth, about half of the non-musicians began tracking spectral information. For amateurs, it was a major third. Professionals were consistently characterized as F_0 classifiers for all intervals, although this became less pronounced as the intervals became closer together (specifically a major second). Seither-Preisler et al.

(2007) acknowledge the possibility that professionals may revert to “spectral” listening if the difference between two tones became smaller than two semitones (a major second), although they never tested this.

Seither-Preisler et al. (2007) outlined three spectral profiles in which their tone pairs were constructed: “(a) low-spectrum tone: 2nd-4th harmonic, high spectrum tone: 5th-10th harmonic, N=17 tone pairs; (b) low-spectrum tone: 3rd-6th harmonic, high-spectrum tone: 7th-14th harmonic, N=17 tone pairs; and (c) low-spectrum tone: 4th-8th harmonic, high-spectrum tone: 9th-18th harmonic, N=16 tone pairs.” Type A tone pairs produced a mean proportion of F_0-based responses at 60.8%. Type B tone pairs resulted in a mean proportion of F_0-based responses at 63.8%, and type C tone pairs demonstrated a mean proportion rate of 55.7%.

The ordering of the tone sequences was controlled in that each trial of the AAT comprised of two tones with different spectral characteristics. Results indicate that the ordering did not make a significant difference. The effect of a falling F_0 and rising spectra were 60.1% for F_0-based responses, and the effect of a rising F_0 and falling spectra in the tone pairs resulted in 60.3% F_0-based responses.

Seither-Preisler et al. (2007) acknowledge the likelihood that only a small percentage of participants responded consistently. Two additional conditions were developed in order to account for guessing amongst participants. The parameters added were defined as “inconsistently categorized tone pairs” and the *“percentage of inhomogeneous judgments”* (Seither-Preisler et al., 2007). Scores for “inconsistently categorized tone pairs” ranged between 0% and 100%, while inhomogeneous scores were limited to a score of 50% (the value appropriate when taking guessing between two modes into consideration).The addition of these parameters presented the ability to check for inconsistency in participant responses, allowing researchers to “identify” which participants were guessing. As a result, Seither-Preisler et al. (2007) reassessed their statistical analyses.

As the AAT was completed twice, Seither-Preisler et al. (2007) were able to track the accuracy of responses for each of the four judgments (50 tone sequences in both orders and the AAT test being performed twice). According to Seither-Preisler et al. (2007), the responses of participants should not have differed for those same four judgments, regardless of when they were presented the sequence. Inconsistencies in responses were assumed to be due to guessing or inconsistency in committing to a “perceptual mode” (either “spectral” or “missing- F_0 classification”). As described by Seither-Preisler et al. (2007), “for spectral classifiers, this is the percentage for F_0-based judgments, whereas for missing-F_0 classifiers, it is the percentage of spectral judgments.”

In order to fully interpret these new parameters, Seither-Preisler et al. (2007) completed thorough Monte Carlo simulations, which allowed the researchers to outline all possible outcomes and the associated probabilities affiliated with decision-making (Palisade, 2019). In this simulation, Seither-Preisler et al. (2007) investigated 100,000 participants, evaluating the virtual participants in an identical manner to how the responses of the real participants were analyzed. The results of the simulation allowed researchers to eliminate participants with “high guessing probability,” conforming with maximum likelihood estimation (MLE). The use of MLE allows researchers to estimate the mean and variance of a large sample while only knowing the data of a small sample (Aldrich, 1997).

The 79 participants were represented on an abscissa with a single data point representing a single participant. Researchers determined the guessing of participants by the clustering of data points. Data points fell within the cluster of "guessing" were immediately removed from the analyses. In total, the responses of 56 participants were analyzed (16 non-musicians, 22 amateurs, and 18 professionals). Originally, there were 30 non-musicians, 31 amateurs, and 18 professionals.

Although half of the non-musicians and a third of the amateurs were excluded from this analysis, musical competence was still highly significant in regard to accuracy in performance. With the original sample, an increase in the proportion of missing-F_0 classifiers mirrored an increase in musical competency; an effect that was also found in the refined sample. In this new sample, 37.5% of non-musicians, 73% of amateurs, and 89% of professionals made their assessments of the sequences based on F_0-pitch cues. This effect, as portrayed with a Mann- Whitney *U* test, was due to the difference of musical experience between non-musicians and the rest of the participant pool. Authors also noted that the instrument played by the musicians did not influence their results; that is, the ability to control pitch (of their instrument) did not influence their accuracy.

2.5 DISCUSSION

Seither-Preisler et al. (2007) noted that the data supported their hypothesis, specifying that the observed differences in performance among professionals, amateurs, and non-musicians were a result of true perceptual differences in relation to musical ability. They state three hypotheses to explain these results.

2.5.1 Hypothesis 1

Hypothesis 1, as noted by the authors, is the most likely explanation of the results. This hypothesis states that playing an instrument (specifically, the enhanced auditory capabilities that come with tuning an instrument) results in an enhanced neural representation of pitch (specifically, the fundamental pitch of complex tones) in the pitch processing areas of the auditory systems. As a result of this cortical plasticity, there would be a significant increase in accuracy in a pitch discrimination task. This hypothesis is further supported by a series of studies using neuroimaging techniques portraying the influence of cortical pitch processing (Krumbholz, Patterson, Seither-Preisler, Lammertmann, & Lütkenhöner, 2003; Penagos, Melcher, & Oxenham, 2004; Griffiths, Buchel, Frackowiak, & Patterson, 1998). Other researchers have developed a model showing that an enhanced neural representation of pitch allows listeners to use learned pitch templates (acquired by listening to voiced speech in early developmental years) to fill in missing information when stimuli were constructed without the F_0 (Terhardt, Stoll, & Seewann, 1982). Another set of researchers proposed a hierarchical auditory model in which the spectral and tem- poral profiles combine to present a single auditory stream (Patterson et al., 1992).Seither-Preisler et al. (2007) cite more studies further emphasizing their theory that increased accuracy in pitch perception tasks with tones constructed without their F_0's is due to cortical neuroplasticity in the auditory regions of the brain. However, Seither-Preisler et al. (2007) do not state the extent to which an instrument should be played (amateur or professional playing, for example).

2.5.2 Hypothesis 2

Seither-Preisler et al. (2007) offer a second hypothesis regarding the noted superior accuracy of professional musicians in pitch perception tasks in which tones are constructed without the F_0. They state that the observed differences in accuracy may be due to ingrained differences between musicians and non-musicians; specifically, a higher sensitivity of the fundamental pitch in complex tones. However, authors do note the need for a longitudinal study (following the effects of musical practice from early childhood to adulthood) to truly assess the influence of genetics on musical aptitude, and specifically on the possible congenital influences regarding the F_0 and complex tones.

Instead of stating the differences are truly due to genetics, authors instead offer another hypothesis; early formative training helping shape the auditory systems of professionals, resulting in a more accurate individual AAT score. They go on to dismiss this claim, however, stating that the onset of musical activity and the first type of instrument played did not have an influence on AAT scores.

2.5.3 Hypothesis 3

As noted in the introduction, pitch occurs on a spectrum much like colour. In Western music, this spectrum is divided into octaves containing 12 semitones. In this study, all the F_0-intervals were derived from this 12 semitone scale with irregular intervals for the spectral series. Professional musicians are at an advantage, especially if they have studied Western style intervals, in that a majority of their music training has them focusing on interval changes (Hannon & Trainor, 2007). Therefore, they may have been more focused on the F_0-intervals as those tones were more musically relevant to what the professionals had spent years training to pay attention to.

2.6 CONCLUSION

Seither-Preisler et al. (2007) concluded that the differences in perception were due to the effects musical expertise had on pitch perception. Specifically, they stated the results were due to the role neuroplasticity and genetics played in musical competency. Professionals, for the most part, were able to track the implied F_0. A majority of the non-musicians tracked the overtone series, and amateurs tracked either the F_0 or the overtones.

2.7 CRITICAL ANALYSIS OF SEITHER-PREISLER ET AL. (2007)

Although this current thesis is based on Seither-Preisler et al. (2007), there are major flaws in study design. First and foremost, it would be nearly impossible to truly assess if the participants were tracking the implied F_0 or the overtones. The AAT consisted of the F_0 and overtones always being in opposition (that is, moving in opposite directions to each other). Including the opposition of the tones in a perceptual task such as this is important, but it is also important to include the tones moving in consonance (in the same direction). In the natural world, F_0's and overtones are always moving in consonance, so it would be well-worthwhile to include this spectral series in the experimental design.

There are both benefits and flaws to the design of the AAT. As tones without the F_0 do not exist in nature, the only way to test this phenomenon is to artificially construct them. As a result, they sound horribly unpleasant and unnatural. This could be a potential confound in an auditory perception task. Secondly, the metaphor of "up" and "down" indicating the "direction" of the tones may be meaningless to individuals who have no experience with such musical terminology; non-musicians may not even fully understand the metaphor of "up" or "down" and as a result not be able to respond accurately.

Professional musicians, on the other hand, base their career on the ability to listen to, identify differences, and overall evaluate pitch. For example, members in an orchestra must tune to each other in order to sound like a cohesive unit, and improvising jazz artists must continually be aware of the changing keys in a chart and adjust accordingly. In Western tradition, when musicians are learning how to tune to those around them, they are often encouraged to hum or sing the pitch that they are trying to match, and as their expertise develops, this humming or singing becomes an internal mechanism that has been fine-tuned over years of practice. The AAT failed to measure the possible use of vocal strategy developed by professional musicians, as they may have implicitly used their own mechanisms (based on musical training) to evaluate pitch. Non-musicians and amateurs likely would not have experienced this training, or in the very least not have developed such an expertise when it comes to evaluating pitch, and this likely was a confound that influenced the accuracy of their performance.

There is an element of learning in this task as well. As mentioned in the previous chapter, individuals can learn to identify pitches over time, either absolutely or relatively. Improved accuracy can be seen throughout one session of a pitch perception task for a variety of reasons, including repeated exposure to the stimuli. In one study, Ben-Haim, Eitan, and Chajut (2014) found that the frequent use of the same pitch in a pitch perception task altered the accuracy rates, simply due to the mere exposure effect. They also stated that a larger part of the population than previously thought likely had a substantial memory for absolute pitch representation. Although not exclusively a pitch perception study, Loui and Wessel (2008) discovered participants were able to apply the knowledge of previously encountered (and recognized) musical stimuli to new melodies.

Seither-Preisler et al. (2007) state that the differences in accuracy among the three participant groups are due to learning induced changes in brain structure that coincide with musical training. The other hypothesis provided state that the differences are due to either genetics or neuroplasticity, and the third hypothesis states that the differences in response accuracy are due to attention on melodic pitch contours. Regardless, the general consensus is that professional musicians, through their musical training, have enhanced neurological capabilities that allow them to perform more accurately in a pitch perception task with tones constructed without the F_0. However, all Western speaking individuals have some accuracy in identifying pitch changes because most Western languages use pitch to convey meaning. Musicians are trained to focus on interval differences between pitch, but neither musicians nor non-musicians would have any real-world experience with tones constructed without the F_0, as they do not occur in nature. Therefore, it would seem likely that non-musicians' expertise of pitch (in regard to speech) and musicians' expertise of pitch (in regard to musical training) would be confounding variables in this type of task. In order to claim that musicians were inherently different, a series of longitudinal studies would need to be conducted. To my knowledge, such studies have not been started, especially in relation to identifying pitches constructed without the F_0, leaving the possibility of future studies being conducted in this domain.

CHAPTER THREE: RESEARCH FOLLOWING SEITHER-PREISLER ET AL. (2007)

Previous research has shown that non-musicians could accurately identify small pitch changes when the fundamental frequency was present in the pitches (Tervaniemi et al., 2004). The next step in this research was to present pitches constructed without the fundamental to participants in a series of tone pairs. Tone pairs consist of two different tones that are presented to participants. The first tone is often considered the "reference" tone, which is then followed by a "test" tone. The two tones will often fall within a range (for example, two tones within the span of an octave). In a pitch discrimination task, participants are asked to differentiate between the tones. Discrimination tasks are the most common way to conduct a pitch perception task. That is, participants are asked to choose whether the test tone is higher or lower than the reference tone. Oftentimes, participants are required to reproduce (usually by singing or humming) the tones presented to them (Welch, 1979; Tervaniemi et al., 2004). As previously stated, overtones are simply integer multiples of the fundamental. Therefore, the pitch of a tone will remain somewhat constant, even after the fundamental is removed (Oxenham, 2012). It has been of interest to researchers how constant and easily perceivable the remaining overtones are to participants (both musicians and non-musicians).

3.1 REPLICATING SEITHER-PREISLER ET AL. (2007): GRANZOW (2010)

Granzow (2010) conducted a series of experiments based on the conclusions of Seither-Preisler et al. (2007). In the first of these, tones were generated to replicate the stimuli of the AAT task used by Seither-Preisler et al. (2007). Participants were recruited from the University of Lethbridge and presented with the pitch perception tasks. They were then presented with a post-test interview classifying them as musical experts, musical amateurs, and non-musicians. Musicians were identified as individuals who have had over five years of formal musical training and who were still regularly performing ($n = 4$). Amateur musicians were classified as individuals with one to five years of musical training ($n = 4$) and non-musicians were defined as individuals with no musical training or participation ($n = 7$), including individuals who played 1-2 years in school (to fill an elective or for curriculum requirements). Seither-Preisler et al. (2007) defined musicians differently, stating the need for classical conservatory education and inquiring about the frequency of musical practice.

The stimuli used by Granzow (2010) were constructed in a way that mimicked Seither-Preisler et al. (2007). Sinusoidal waves for each tone were summed, and those summations corresponded to the harmonic series (integer multiples) of the given F_0. The intensity was then normalized. The implied F_0 and associated overtones always moved in opposition, resembling the stimuli of Seither-Preisler et al. (2007). The lowest implied F_0 used by Granzow (2010) was 100 Hz and the highest implied F_0 reached 400 Hz. As in Seither-Preisler et al. (2007), half-steps associated with the Western major musical were used (semitones 2, 4, 5, 7, or 9). The tones were split into groups of three; group A consisted of the higher of the F_0's being implied by harmonics 2-4 and the lower of the F_0's being implied by harmonics 5-10. In group B, the higher of the implied F_0's were portrayed by harmonics 3-6 and the lower harmonics 7-14. In group C, harmonics 4-8 implied the higher F_0's and harmonics 9-18 implied the lower F_0's. In total, there were 3 spectral series (groups A, B, and C), 5 semitone difference categories (semitones 2, 4, 5, 7, or 9), and 2 orders for each tone pair (either ascending or descending). In total, there were 210 randomized trials for each participant. Participants were presented with a tone-pair and asked to give a two-alternative response choice. They were instructed to indicate whether the tones went “up” (by pressing the “?” key) or “down” (by pressing the “Z” key). As Granzow (2010) was replicating Seither-Preisler et al. (2007), the instructions suggesting participants hum or sing the tones were included.

As in Seither-Preisler et al. (2007), Granzow (2010) found that musicians were the most accurate in the pitch discrimination task, followed by amateurs and then non-musicians. Granzow (2010) noted a main effect of semitone differences, which was also recognized in Seither-Preisler et al. (2007); larger semitone differences resulted in more accurate responses, with non-musician performance falling to chance at the smallest semitone differences. Spectral profile pairings had no influence on performance. Granzow (2010), in contrast to Seither-Preisler et al. (2007), found a significant interaction between expertise and semitone differences; amateurs and non-musicians both consistently tracked the implied F_0 as the semitone differences increased. Musician performance was not influenced by semitone differences. The results regarding expertise and semitone accuracy were reflected in the response patterns of the AAT used in Seither-Preisler et al. (2007).

Granzow (2010) offers a different explanation regarding the poorer performance of non- musicians, especially with tone pairs containing smaller semitone differences. He states the metaphor of tones going “up,” “down,” and “pitch height” may be subjective and actually hinder their performance, as they may not have the experience with musical vocabulary. Previous research has indeed shown that linguistic categories do have an influence on perception (Stapel & Semin, 2007). Musical expertise would also account for the association of musical metaphors and playing, an experience the amateurs or non-musicians likely would not have.

In order to account for the influence of musical metaphors, Granzow (2010) tested a group of non-musicians (*n* = 13) and gave them feedback regarding how well they tracked the implied F_0 after every trial. The materials and procedure for this smaller group were identical as before, with the exception of receiving the feedback ("correct" or "incorrect" were displayed on the screen following each trial). As in the first group, accuracy was influenced by the difference in semitone intervals (higher accuracy for larger semitone differences). However, the feedback group performed significantly better in the trials where the semitone interval differences were smaller, unlike the non-musicians in the first study. Overall, the feedback (regarding how well participants were tracking the implied F_0) slightly improved the performance of non-musicians, but their performance was not anywhere near the musicians' level of accuracy.

The AAT designed by Seither-Preisler et al. (2007) allowed for variability in the presentation of the tones. The reference tone (always presented first) was subject to change, as was the test tone (presented second). There was also variability of the nature of the overtones; they were either rising or falling in relation to their opposite implied F_0's. For simplicity's sake (and for the nature of the experimental design), providing a consistent reference tone may allow experimenters to better measure the abilities of the participants by allowing them to solely focus on the test tone.

In the next experiment, Granzow (2010) used a consistent reference tone followed by a variable test tone. The pitch of the test tone was always falling or rising (in relation to the reference tone) by 2, 4, 6, or 8 semitones. The reference tone (and associated implied F_0) always corresponded to piano key number 40 (middle C, or C4). The remaining overtones for the reference tone were structured from the fifth overtone through the to the tenth overtone.

Three spectral series were used in order to parse out the sources of responding by the participants; whether or not they were tracking the implied F_0 or the overtones. The spectral series of the second tones were labelled as Low, Middle, and High. The Low spectral series consisted of 8 tones constructed from the low octave of overtones 2 through 4. As a result, even if the implied F_0's of the test tone was "higher" than the reference tone, the actual overtones would be lower in frequency. In other words, the implied F_0 of the test tone and associated overtones were consonant below middle C but in opposition above middle C, much like the spectral series in the original AAT designed by Seither-Preisler et al. (2007).

The Middle spectral series consisted of 8 tones constructed from the middle octave of overtones 5 through 10. As with the Low spectral series, the same tones were used as those generating the reference tone. Unlike the Low spectral series, both the implied F_0's and the frequencies of the overtones always moved in the same direction. The High spectral series similar to the Low spectral series; the direction of the implied F_0's moved in consonant with overtones above middle C, but in opposition below middle C, as in the AAT. The High spectral series consisted of 8 tones constructed from the high octave of overtones 9 through 18. if the implied F_0's of the test tone was "lower" than the reference tone, the actual overtones would be higher in frequency.

These tones were not only constructed without the fundamental frequency, but also were designed in such a way that the implied fundamental and the overtones were either in opposition or moving in the same direction (by definition of the spectral series). If participants were tracking the implied $F0$, then they would perform similarly regardless of which spectral series they were presented with, as the semitone differences would not affect their accuracy. If they were tracking the overtones, they would perform better on the high end of the high stimuli and the low end of the low stimuli, but their best performance would be with the Middle stimuli.

The post-test criteria were used to classify musicians and non-musicians; 8 musicians and 14 non-musicians participated in this experiment. The three spectral series and eight test tones resulted in 24 randomized trials per the 4 repeated blocks, resulting in a total of 384 trials. Participants were instructed to identify the direction of the test tone (in relation to the reference tone) as quickly (but as accurately) as possible.

Musicians, surprisingly, did not perform significantly better than non-musicians at tracking the implied $F0$. There was no effect of musical expertise. There was a significant effect of semi- tones; as in the previous experiment of Granzow (2010), test tones with implied $F0$'s generating the smallest difference in relation to the reference tones were the hardest for participants to identify. There was also a main effect of spectral series; the Middle spectral series generated the most accurate performance, followed by the High spectral series, and the Low spectral series generating the least accurate responses.
The most significant effect was the interaction between the semitones and spectral series. The tones that moved in consonance generated more accurate results than those with incongruent movement (for example, the Low spectral series with implied $F0$'s that were

higher than the reference tone). The results of the incongruent tones and semitone differences generated results similar to those in the first experiment of Granzow (2010) for non-musicians. Granzow (2010) noted the marked difference in accuracy of the musicians between the two experiments; musicians performed similarly to non-musicians in this experiment.

The next experiment of Granzow (2010) was identical to the pervious one, except instead of being asked to respond quickly, participants were asked to reproduce (hum or sing) the two tones on each trial prior to making their judgment, although their vocal reproductions were not recorded. If the musicians were humming or singing in the previous experiments (and if this vocal strategy produced a better performance) then perhaps instructing non-musicians and amateurs to do the same would improve their performance.

Seven musicians and seven non-musicians completed this experiment. The stimuli and procedure were identical to those used in the previous experiment except for the instructions regarding speed; instead, participants were instructed to hum or sing both tones prior to responding. In contrast to the previous experiments, expertise had a significant effect on accuracy. In this experiment, musicians were more consistent in tracking the implied F_0's (similar to results in the first experiment of Granzow (2010) and results noted by Seither-Preisler et al. (2007)). There was a main effect of spectral series, with tones in the Middle spectral series being tracked more accurately than those in the High or Low spectral series. Again, as in the previous experiment, there was a main effect of semitone difference between the two tones and a significant interaction between spectral series and semitone differences.

Overall, musicians were unaffected by spectral series, semitone differences, or interactions between the two. The instruction to reproduce vocally the tones aided in the accuracy of musician performance. In contrast, non-musicians were significantly affected by spectral series, semitone differences, and the interactions between the two. Having the non-musicians vocally reproduce the tones did not result in any significant differences in performance.

Unlike Seither-Preisler et al. (2007), Granzow (2010) was able to begin to parse out what aspects of the tones the participants were able to truly track (due to the nature of the tones moving in either opposition or in consonance). Both the spectral series (also known as the overtone series) and the distance between the reference tone and the test tone affected the performance of non-musicians but not musicians, unless musicians were asked to respond quickly. The poorer performance of musicians in second experiment was attributed to the speeded condition of the task, preventing musicians from doing what they have been doing naturally (and possibly implicitly) in previous experiments (Granzow, 2010) and (Seither-Preisler et al., 2007)): evaluate pitch (as discussed in the previous chapter). As a result, Granzow (2010) conducted a third experiment explicitly instructing participants to vocally reproduce the tone pairs in each trial before responding. Results from this experiment indicated that vocal strategies were being used by musicians, as their performance accuracy significantly increased. Results also indicated that instructing non-musicians to reproduce vocally the tone pairs did not result in significantly improved performance, although there is still the possibility that training the non-musicians to respond similarly to musicians would result in improved accuracy over time. They did note one non-musician who acknowledged that vocally reproducing the tones aided their performance regarding the direction of the tone pairs.

3.2 A STEP FURTHER: REIMCHEN (2011)

The next step was to identify the role musical expertise played in performance accuracy. Specifically, the role musical expertise had regarding the increased accuracy portrayed by musicians after being instructed to vocally reproduce the tone pairs. Granzow (2010) demonstrated that performance accuracy in musicians was dependent on their ability to vocally reproduce tones (regardless of the nature of the overtones), although this vocal strategy did not improve the performance of non-musicians. Previous research has shown that vocal strategies can be helpful in an auditory perception task, especially when the tones constructed without the F_0 and if the participants had some level of musical training (Thurlow, 1963; Thurlow & Hartman, 1959). Reimchen (2011) attempted to identify the role musical expertise played on sung- direction accuracy for both musicians and if vocal strategies could help improve accuracy of non-musicians. The vocal reproductions of the participants were recorded and analyzed along with their indicated direction of the tones.

The pitch contour paradigm used in this experiment was designed by Granzow (2010), which was influenced by the AAT paradigm proposed by Seither-Preisler et al. (2007). Again, all tones in this experiment were constructed without the F_0. Ten musicians, five amateurs, and ten non-musicians participated in this experiment. Musicians, amateurs, and non-musicians were identified by the same post-test questionnaire used by Granzow (2010). This experiment consisted of both a production task (vocally reproducing the tone pairs) and a perception task [identifying whether the test tone (presented second) was higher or lower than the reference tone (presented first)].

The tones and spectral series used in this experiment were the same as those used in the last two experiments of (Granzow, 2010). The tone pairs occurred in four randomized blocks, each consisting of 24 trials. The blocks were repeated 4 times; in total, participants were presented with 384 trials. Participants were instructed to reproduce vocally (hum or sing) each tone pair before indicating whether the test tone was higher or lower than the reference tone. Vocal reproduction responses were recorded with a microphone and their indication of pitch height was recorded via computer keyboard presses (“z” or “spacebar” for lower and “/” or “6” for higher).

Given the smaller sample size of amateurs, they were excluded from analyses for this experiment. Responses were recorded in two ways; vocal recordings (recorded on a microphone) and explicit computer keyboard presses indicating the direction of the tone pairs. If participants were tracking the implied F_0’s, there should be no observed difference on which spectral series they performed better on. In contrast, if the participants were tracking the overtones, they should perform less accurately on spectral series in which the overtones and implied F_0’s were incongruent. As with Granzow (2010), participants were expected to perform best when the semitone difference was the largest, with poorest performance occurring when the semitone difference was smaller.

Vocal reproductions were analyzed via PRAAT software (Boersma & Weenink, 2011). The frequencies of the sung fundamentals were extracted in order to identify whether or not the participants sung in the correct direction and to identify the accuracy of their singing. Regardless of explicit direction (indicated by computer keyboard presses), non-musicians were generally able to vocalize the correct direction of the tone pairs, even if they indicated the opposite direction via computer key-presses. There was no effect of frequency (pitch vocalized), spectral series, or any interaction of the two in regard to vocal reproduction recorded by non-musicians. However, as seen in previous research, non-musicians were influenced by interval size when responding by computer key presses. As in the previous studies discussed, non-musician accuracy was reduced when there were small semitone differences between the tone pairs, and there was a significant interaction between spectral series and semitone differences. However, Reimchen (2011) noted the benefit of vocally reproducing the tones, as the poor performance of non-musicians in this study was, overall, more accurate than the performance of non-musicians in experiments where they were not asked to vocally reproduce the tone pairs.

Musicians were accurate in their key-presses (as seen in the previous studies) and were not influenced by frequency, spectral series, or any interaction between the two. Unlike the dissociation seen in the non-musicians regarding the accuracy of vocal reproduction and key- presses, musicians were accurate in both domains. Musicians were also unaffected by semitone differences. Overall, musicians were more accurate than non-musicians in both key-presses and sung responses.

Sung tones were also analyzed to identify how accurate the vocal reproductions were in regard to sung pitch. Musicians were highly accurate in sung pitch, regardless of frequency and spectral series. Non-musicians, on the other hand, were highly variable for both frequency and spectral series, likely due to difficulties integrating production, memory, and sensorimotor information (Pfordresher & Brown, 2007). The expertise of musicians in their ability to evaluate pitch may be so deeply engrained that they may not even be aware of it, as their capacity for pitch processing may exist in an experience-dependent capacity (Koelsch & Siebel, 2005).

CHAPTER FOUR: OVERVIEW OF THE EXPERIMENTS

4.1 EXPERIMENTAL DESIGN

Having participants (both musicians and non-musicians) vocally reproduce the tones affected the accuracy of their performance (Granzow, 2010; Reimchen, 2011). However, as noted by Reimchen (2011), non-musicians were able to reproduce vocally the tones in the correct direction, but there was still a dissociation between their vocal production, pitch memory, and indication of pitch direction via key-presses. Musicians, on the other hand, were more accurate in their vocal reproductions as well as being better able to identify (via computer key-presses) changes in pitch. Understanding the processes by which musicians complete these tasks may allow us to provide support for non-musicians, allowing them to obtain musician-level accuracy. What may differ between musicians and non-musicians, with respect to their ability to discriminate among pitch changes, is the proper use of tonal memory. In order for musicians to successfully tune their instruments (or themselves), they need to be explicitly aware of the note they are trying to match, and consistently and accurately hum it or think about it (Welch, 1979). Over the length of their musical training, this habit becomes second nature to them, as they are able to successfully attend to the note they are singing or humming. Therefore, it seems that training (or repeated exposure to stimuli) will help improve non-musician performance, as well as attending to various aspects of vocal reproduction.

In the previous experiments, participants were required to indicate their responses via computer keyboard presses (one key indicating that the tones went up, and another indicating the tones went down) (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). In the last study examined in the previous chapter, participants were asked to vocally reproduce the tones before indicating their responses on a computer keyboard (Reimchen, 2011). In this thesis, we included the traditional paradigm (computer key-presses indicating responses) as well as the vocal reproduction and key-press paradigm used by Reimchen (2011), but we also included two more conditions outlined below. The inclusion of these conditions were to assess whether or not different responses to a pitch discrimination task aided in accuracy (and to what extent), as well as assessing whether or not repeated exposure to the tones improved accuracy. In particular, we were interested in what aspects of vocal reproduction aided in performance accuracy.

In the previous studies mentioned (with the exception of Reimchen (2011)), the standard paradigm has been leaving the participants to their own devices in regard to responding to the pitch changes. In this thesis, this condition has been labelled *no instruction*, as participants were not given any instruction regarding vocal reproduction of tones, instead just being instructed to indicate the direction of the tones by responding via computer key-presses. The other three conditions were included in order assess whether the non-musicians were impacted by the tones themselves or by the responses they were required to produce.

The *vocal reproduction* condition follows the paradigm used by Reimchen (2011). Vocal reproduction (or auditory feedback) of sounds in the environment helps in our perception of the world around us, as we can modify our reproductions in order to hear the world in a clearer manner (Jones & Keough, 2008). If sounds in our environment have been manipulated by external conditions (for example, tones constructed without the $F0$ or a masked noise), auditory feedback can help modify our perception of such sounds by forcing us to modify our reproductions in order to make sense of our auditory world

(Jones & Keough, 2008). In other words, vocally reproducing unnatural tones may allow participants to better attend to the tones they are being presented.

Reimchen (2011) noted an increase in performance accuracy (for both musicians and non- musicians) after instructing participants to vocally reproduce the tones. By vocally reproducing tones, two types of information are generated that may be helpful in making the decision; first, by singing (and matching) the presented tones, participants are automatically reintroducing the F_0, thereby increasing the accuracy (or at least, making the task easier) in identifying tonal direction, as tones naturally occur with the F_0. Second, by physically singing the tones, participants induce laryngeal shifts, either straining to hit high notes or compressing their larynx to hit lower notes (Sonninen, Hurme, & Vilkman, 1992). The *vocal reproduction* condition will assess the influence of both types of information (the reintroduction of the F_0 and the laryngeal shifts) in performance accuracy.

The *slider* condition was included to eliminate the vocal production of tones (and the subsequent addition of the F_0), while maintaining a noticeable physiological response. This condition was included as a measure to identify whether or not vocally replicating the tones (and adding the F_0) aided in participant performance accuracy, or whether attending to a physical reaction (such as laryngeal movement) was a source of the increased accuracy. The slider was used to attempt to mimic subtle laryngeal movement that occurs when an individual sings, especially in non-musicians. When a singer strains to sing high notes, the muscles in the throat tense and push the larynx upward, in contrast to the lower placement that occurs when the throat is relaxed and low notes are sung (Sonninen et al., 1992). Consequently, participants have two separate opportunities to identify the direction of the tones; by attending to the vocalizations they produce, or by detecting subtle muscular (laryngeal) movements that take place when singing occurs.

The *speeded* condition was included as a means of eliminating all responses associated with vocal reproduction (the reintroduction of the F_0 as well as muscular feedback obtained from laryngeal movement). In this condition, participants were required to respond to the pitches as quickly as possible, thereby eliminating any possibility of laryngeal movement or humming.

4.2 CONDITIONS

Participants were presented with four conditions, all requiring different responses to a pitch perception task with tones constructed without the F_0. Participants were brought into the lab and, after obtaining consent, were instructed to sit at a lab computer and read the instructions presented on the screen. The instructions differed based on the condition.

4.2.1 No Instruction

In the *no instruction* condition, participants were given no instructions regarding vocally reproducing the tones or completing the task as fast as possible. They were simply instructed to identify the direction of the tones ("/" for when they thought the tones went up and "z" for when they thought the tones went down). It was of interest to see how responses would (or would not) differ when participants were left to their own devices in comparison to performance accuracy in other conditions. This has been the standard paradigm for both Seither-Preisler et al. (2007) and Granzow (2010).

4.2.2 Vocal Reproduction

The *vocal reproduction* condition most mirrored the experiments conducted by Reimchen (2011). The two tones were presented to participants and they were asked to sing or hum both tones before indicating their response. Their vocal reproductions automatically contained the F_0 of the tones, assuming they were relatively close to matching the Hz of the implied F_0. They then recorded their responses on a computer keyboard, again using "/" to indicate the tones went up and "z" to indicate the tones went down.

4.2.3 Slider

The *slider* condition was unique, in the sense that responses were recorded via a vertical slider on the screen. Participants were asked to move the slider to the extent to which they thought the test tone (presented second) differed from the reference tone (presented first).The only predetermining parameters on the slider were whether or not the test tone went up or down; it was entirely up to the participant to decide the extent to which the test tone differed from the reference tone (*how* much higher the test tone was in comparison to the reference tone).

4.2.4 Speed

In the *speeded* condition, participants were instructed to complete the task as quickly as possible, thereby eliminating any opportunity to hum or think of the implied fundamental. Their responses were recorded on a computer keyboard, with "/" indicating that the tones went up and "z" indicating the tones went down.

4.3 PARTICIPANTS

Participants were recruited from the undergraduate population at the University of Lethbridge. Participants were recruited from both undergraduate psychology courses and undergraduate music courses. Psychology students were mainly recruited from SONA, the University of Lethbridge's online research recruitment website, and they received one course credit for participating in the study. Participants were classified via gender, age, and musicianship. Those with recent musical training and who currently played an instrument or sang were considered musicians.

Recent musical training was defined as receiving instruction (either private or in a larger class) within the last 5 years, regardless of the quantity of the lessons received. Training was defined as practicing or learning music theory and/or how to play an instrument or sing within the last 5 years. In order to be defined as a musician, participants were required to have played an instrument or to have actively engaged in singing within the last 5 years. Playing an instrument was defined by recreational, professional, or academic engagement in the physical practice of playing an instrument (for example, a piano or trumpet). The ability to read music was not included as a defining feature of a musician; for example, Elvis Presley could not read music, but he still earned the title of the "King of Rock and Roll" (Patil, 2018). As musicians can still be highly proficient in their musical specialty, the ability to read sheet music was not included as an identifying parameter of musicianship. In previous research, pitch discrimination abilities did not differ between instrumentalists and vocalists, but instead between musicians and non-musicians (Nikjeh, Lister, & Frisch, 2009). Therefore, both vocalists and instrumentalists were collapsed into the group "musician."

Individual vocal range is diverse, and especially varies across gender. On average, male and female voices are about an octave apart (Titze, 2004). Males tend to have trouble matching pitches (physically reproducing the tones) in higher ranges, and females have trouble matching pitches in lower ranges. However, it is natural for humans to sing octaves with one another; the acoustic relationship of an octave is naturally found within harmonics. Therefore, in the following studies, vocal range is not something to be seriously considered. Participants will naturally transpose the tones heard to a comfortable octave, and as octaves are harmonically similar, transpositions will not affect the data.

4.3.1 Modified Latin Square Design

As this was a within-subject design, we did not want any carryover effects (in which being tested in one condition influenced the responses in another). For example, if a participant began the study in the *vocal reproduction* condition, they may be more inclined to hum or sing (or even mentally hum and sing) the tones in the subsequent conditions, thereby influencing their performance accuracy.

The four conditions (*no instruction, vocal reproduction, slider,* and *speeded*) preceded and followed each other equally often across participants. For example, if the *no instruction* condition was labelled "a," the *vocal reproduction* condition labelled "b," the *slider* condition labelled "c," and the *speeded* condition labelled "d," the four counterbalancing conditions would be as followed: abcd, cdab, dcba, and badc. Each letter (condition) occurred once in each position, with each letter (condition) preceding and following the other three conditions equally often.

Responses were subjected to a mixed ANOVA (testing for differences between two or more independent groups with participants being presented with repeated measures).Analyses of proportion correct (pc) of discrimination accuracy were conducted in two ways; including performance of a participant across all four conditions (as a within-subjects design), and treating the conditions as a between-subject design by only analyzing the initial (first) condition presented to participants. This was done as an attempt to account for any possible carryover effects (particularly, humming or singing in multiple conditions regardless of instruction).

Vocal reproductions were recorded and analyzed via Praat, a software program that allows for the pitch analysis of vocal recordings (Boersma & Weenink, 2011). There were three analyses conducted with participant vocal reproductions; analyzing potential differences between key-press and sung direction accuracy, analyzing the accuracy of the participant in reproducing the reference tone, and analyzing the sung accuracy of the test tone.

4.4 OVERTONES AND FREQUENCIES

Tones were structurally similar to those used by Seither-Preisler et al. (2007), Granzow (2010), and Reimchen (2011). Sinusoidal waves for each tone were summed and corresponded to the overtone series (integer multiples) of the given (implied) fundamental. The interval between the highest and lowest overtone was always an octave, with the intensity normalized. Tones were presented as pairs; the reference tone (presented first) remained constant across all trials (the tone itself was middle C, or 261.63 Hz) and was constructed of overtones 5 through 10. This was followed by the test tone, which ranged from 2 to 8 semitones above or below the reference tone. Participants were asked to identify the degree to which the test tone differed from the reference tone in a variety of tasks for both experiments.

4.5 SPECTRAL SERIES

Test tones occurred within one of three spectral series; Low, Middle, and High, mirroring the makeup of stimuli observed in Granzow (2010) and Reimchen (2011). The stimuli used mirrored the AAT developed by Seither-Preisler et al.

(2007), although slightly altered. Seither-Preisler et al. (2007) conducted their studies with tones in which the overtones and implied F_0's always moved in opposition (Low and High spectral series), and the reference tone presented to participants differed based on spectral series. Granzow (2010), following the lead of Seither-Preisler et al. (2007), included the spectral profiles with the implied F_0's and overtones moving in opposition, but also included a spectral profile in which the implied F_0's and overtones moved in concert; the Middle spectral series. This was included as a control by being able to measure which aspects of pitch participants attended to in order to complete this task. He also presented the same reference tone to all participants, regardless of spectral series (middle C). Reimchen (2011) used the three spectral profiles outlined in Granzow (2010), instead altering the responses participants were required to elicit (by including a paradigm requiring participants to vocally reproduce the tones).

Based on previous research conducted [(Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011)], we predicted that if participants were tracking the implied F_0, then responses would be similar regardless of spectral series and uninfluenced by semitone differences. In contrast, if they were tracking overtone frequencies, then their accuracy would showcase patterns of response accuracy (in regard to semitone differences) in both the Low and High spectral series. In particular, we predicted that both musicians and non-musicians would have the most trouble in the Low spectral series followed by the High series, as the manipulation (having the overtones and implied F_0's move in opposition) could produce conflicting auditory stimuli that had the potential to make the task more difficult. The spectral series were counterbalanced among all four conditions, so that the presentation of one of the spectral series did not influence responses to tones constructed by another spectral series.

4.5.1 Low Tones

Low tones were constructed from the low octave overtone frequencies 2 through 4. As a result, the test tones consisted of overtones with an overall lower frequency than the overtones used to generate the reference tone. Implied F_0's and overtone frequencies moved in concert for test tones below middle C but in opposition when moving above middle C.

4.5.2 Middle Tones

Middle test tones were constructed from the middle octave of overtones (5 through 10), which were identical to the octave of overtones present in the reference tone. As a result, reference tones and test tones always moved in concert (both above and below middle C).

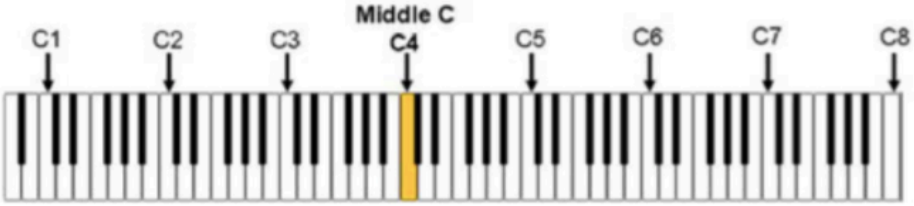

Figure 4.1: Example of a piano keyboard outlining the placement of C4.

4.5.3 High Tones

High tones were constructed from the high octave overtone frequencies 9 through 18. Overtones of the test tones, as a result, were overall higher in frequency than the overtones generated by the reference tone. As a result, the tones in this condition were opposite to those in the Low condition; Implied F_0's and overtone frequencies moved in concert for test tones above middle C but in opposition when moving below middle C. Shown in Figure 4.1 is an example of a piano keyboard outlining the placement of middle C or C4. Shown in Figure 4.2 is a schematic of the experimental design of the spectral series.

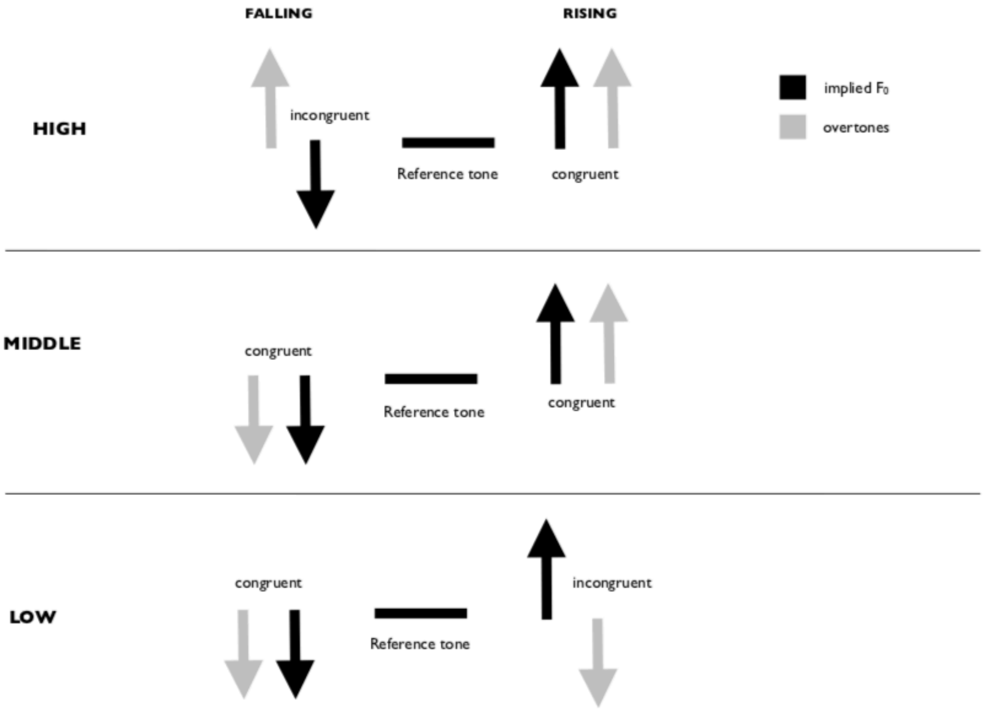

Figure 4.2: schematic of the experimental design of the spectral series for Experiment 1.

Table 4.1: Frequencies (Hz) of the Overtones Used to Imply the Missing Fundamentals of the Tones as a Function of Spectral Series (Low, Middle, and High) and Semitone (Note on the Piano Scale).

		Tones								
	Note	E3	F♯3	A♭3	B♭3	C4	D4	E4	F♯4	A♭4
	Piano Key	32	34	36	38	40	42	44	46	48
	F_0	164.81	185	207.65	233.08	261.63	293.66	329.63	369.99	415.3
Low	2	329.63	369.99	415.3	466.16		587.33	659.26	739.99	830.61
	3	494.44	554.99	622.96	699.25		880.99	988.88	1109.98	1245.91
	4	659.26	739.99	830.61	932.33		1174.66	1318.51	1479.98	1661.22
Medium	5	824.07	924.99	1038.26	1165.41	1308.13	1468.32	1648.14	1849.97	2076.52
	6	988.88	1109.98	1245.91	1398.49	1569.75	1761.99	1977.77	2219.97	2491.83
	7	1153.7	1294.98	1453.57	1631.57	1831.38	2055.65	2307.39	2589.96	2907.13
	8	1318.51	1479.98	1661.22	1864.66	2093	2349.32	2637.02	2959.96	3322.44
	9	1483.32	1664.97	1868.87	2097.74	2354.63	2642.98	2966.65	3329.95	3737.74
	10	1648.14	1849.97	2076.52	2330.82	2616.26	2936.65	3296.28	3699.94	4153.05
High	9	1483.32	1664.97	1868.87	2097.74		2642.98	2966.65	3329.95	3737.74
	10	1648.14	1849.97	2076.52	2330.82		2936.65	3296.28	3699.94	4153.05
	11	1812.95	2034.97	2284.18	2563.9		3230.31	3625.9	4069.94	4568.35
	12	1977.77	2219.97	2491.83	2796.98		3523.98	3955.53	4439.93	4983.66
	13	2142.58	2404.96	2699.48	3030.06		3817.64	4285.16	4809.93	5398.96
	14	2307.39	2589.96	2907.13	3263.15		4111.31	4614.79	5179.92	5814.27
	15	2472.21	2774.96	3114.79	3496.23		4404.97	4944.41	5549.92	6229.57
	16	2637.02	2959.96	3322.44	3729.31		4698.64	5274.04	5919.91	6644.88
	17	2801.83	3144.95	3530.09	3962.39		4992.3	5603.67	6289.91	7060.18
	18	2966.65	3329.95	3737.74	4195.47		5285.97	5933.3	6659.9	7475.48

CHAPTER FIVE: EXPERIMENT 1: RAMPED TONES

5.1 PARTICIPANTS

48 non-musicians and 40 musicians completed this experiment. Participants were seated at a computer monitor and keyboard. They were given headphones and presented with the letter of consent. After indicating their consent, they were then instructed to begin the experiment. Vocal reproductions were recorded via microphone and participants were given instructions on how to adjust the volume of the tones.

5.2 MATERIALS AND PROCEDURE

The tones outlined in the previous chapter were the stimuli in this experiment. The tones were digitized to 8-bit WAV files with a sampling frequency of 44,100 Hz, with a duration of 500 msec. The presentation of the tones followed a Gaussian function. The Gaussian function is described as the "probability density function of the normal distribution (in one dimension)," and by taking on the Fourier transform, the tones themselves will follow the shape of a bell curve (Seither-Preisler et al., 2007). The onset of the tones were ramped, in that they slowly reached peak volume at the peak of the ramp (and therefore presenting the full amount of auditory information), followed by a ramped offset. All the tones used in this experiment were constructed without the F_0, just leaving the overtones.

The conditions were counterbalanced so that each participant started on one of the four, attempting to prevent any one condition from influencing their response in the other conditions. The *no instruction* condition was included as it followed the standard paradigm used in previous studies (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). The *vocal reproduction* also replicated conditions used in previous research, and it was established that vocally reproducing the tones significantly increased performance accuracy (Reimchen, 2011). This thesis attempted to tease apart what aspects of vocal reproduction were important (the re-addition of the F_0 or the shifting of the larynx), by introducing two new conditions; the *slider* condition and the *speeded* condition. The *slider* condition was used to mimic the movement of the larynx (in a somewhat embodied manner) while the *speeded* condition was included in an effort to exclude any manner of vocal reproduction from occurring by having participants respond as quickly as possible. The *no instruction, vocal reproduction,* and *speeded* conditions all required participants to indicate their responses via computer key-presses ("z" for down and "/" for up) following the first set of instructions. The *slider* condition required participants to indicate how much higher or lower they thought the test tone was from the reference tone via a vertical slider on the computer screen, and the slider responses were then translated into a direction for analyses. The combination of 3 spectral series and 8 possible semitone keys were repeated four times, resulting in 96 trials per condition.

5.3 RESULTS

The experiment resulted in two different measures of performance: (1) judgements of pitch direction as given by the key-press and slider responses, and (2) the actual frequencies produced by participants in the *vocal reproduction* condition, and how those vocal reproductions compared to key-press responses. For the key-press responses (the *no instruction* condition, the *speeded* condition, and the *vocal reproduction* condition), a response scored as concordant if it was consistent with the direction of the two tones; for the *vocal reproduction* condition, a key response was scored as concordant with "/" associating with the "up" direction and "z" associated with the "down" direction of the tones for that trial. These were then summed and converted to proportions of concordant responses for each experimental condition.

5.3.1 Proportion Correct

The proportion correct—concordant with F_0—was subjected to a 4 (Counterbalancing Group) × 2 (Musician Type: musician, non-musician) × 4 (Test Condition: no instruction, sing, slider, speeded) × 3 (Spectral Series: high, low, middle) × 8 (piano Key number: 32, 34, 36, 38, 42, 44, 46, 48) mixed analysis of variance with participants crossing Test Condition, Spectral Series, and Piano Key, but nested within Counterbalancing Group and Musician Type. The Greenhouse-Geisser degrees of freedom correction for sphericity was used for all effects involving within-participant factors. Overall, musicians tracked F_0 significantly more accurately than non-musicians (as seen in Figure 5.1): $F(1,80) = 12.88, MS_e = 0.98, p = 0.001$.

Participants (both musicians and non- musicians were influenced by the spectral series, with both groups performing most accurately with tones constructed in the Middle spectral series: $F(1,147) = 36.84, MS_e = 0.10, p = 0.0001$, $\eta^2 = 0.02$. Both groups of participants were also influenced by the size of the interval between the reference and test tone: $F(2,215) = 52.31, MS_e = 0.19, p = 0.001, \eta^2 = 0.07$. Musicians and non-musicians were also influenced by a spectral series by piano key interaction: $F(2,204) = 41.16, MS_e = 0.34, p = 0.001, \eta^2 = 0.09$.

The next analysis only analysed only the first condition each participant completed, as a means to control for any possible carryover effects (the possibility of responses in one condition influencing the responses in another). Overall, there was no significant difference between musician and non-musician response accuracy. However, both musicians and non- musicians were influenced by spectral series [$F(1,146) = 29.06, MS_e = 0.06, p = 0.001, \eta^2 = 0.03$], piano key [$F(3,296) = 25.89, MS_e = 0.08, p = 0.001, \eta^2 = 0.07$], spectral series by piano key [$F(3,273) = 28.46, MS_e = 0.15, p = 0.001, \eta^2 = 0.13$], and condition by piano key [$F(11,296) = 1.84, MS_e = 0.08, p = 0.05, \eta^2 = 0.02$].

5.3.2 Vocal Reproduction

Three analyses were conducted on the vocal reproductions of the participants: (1) how concordant was the vocal production of the tones with the direction of the tone pairs (similar to the key-press response analysis); (2) how accurately did participants reproduce the frequency of the reference tone (middle C, or C4), and (3) how accurately did they reproduce the test tone of each tone pair?

5.3.3 Key Press versus Sung Direction

The fundamental frequencies of the two tones (reference tone and test tone) vocalized in the *vocal reproduction* condition for each participant were isolated using PRAAT (Boersma & Weenink, 2011). Shown in Figure 5.3, the key-press versus sung direction in the vocal reproduction condition of Experiment 1 displayed marked differences between musicians and non-musicians. Musicians were significantly more accurate in their key-press responses compared to non- musicians: $F(1,64) = 4.60, MS_e = 0.43, p =< 0.0001, \eta^2 = 0.02$. sung direction was scored as concordant or not concordant with the direction of the fundamental. Musicians were also more accurate in their vocal reproductions: $F(1,64) = 10.15, MS_e = 0.20, p = 0.0001, \eta^2 = 0.02$.

Overall performance accuracy for both musicians and non-musicians was influenced by spectral series, with participants performing more accurately with tones constructed through the Middle spectral series: $F(1,116) = 13.24, MS_e = 0.05, p = 0.0001, \eta^2 = 0.01$. Participants were also influenced by interval key, with performance accuracy worsening as the intervals between the tones became smaller: $F(3,230) = 9.75, MS_e = 0.08, p = 0.0001, \eta^2 = 0.02$. There was also an interaction between spectral series and interval key: $F(5,382) = 13.62, MS_e = 0.06, p = 0.0001, \eta^2 = 0.04$.

Singing accuracy (for both musicians and non-musicians) was influenced by interval key, with accuracy decreasing as the intervals between the tones became smaller: $F(4,291) = 9.64, MS_e = 0.04, p = 0.0001, \eta^2 = 0.01$. Singing accuracy was also influenced by an interaction between spectral series and interval key: $F(6,438) = 5.53, MS_e = 0.03, p = 0.0001, \eta^2 = 0.01$.

Overall, there was a discord between sung direction and key press: $F(1,64) = 10.15, MS_e = 0.19, p = 0.0001, \eta^2 = 0.02$. That is, participants would reproduce the notes in one direction (say, with the test tone higher compared to the reference tone) but indicate via key-press that the test tone was lower than the reference tone. Overall, participants were more likely to sing the tones in the correct direction and indicate their key-press response in the wrong direction (key: M = 0.89, SD = 0.13 and sung: M = 0.93, SD = 0.10). The discord between key-presses and sung direction were also influenced by interval key: $F(4,291) = 9.64, MS_e = 0.04, p = 0.0001, \eta^2 = 0.001$, spectral series: $F(1,97) = 4.99, MS_e = 0.02, p = 0.0001, \eta^2 = 0.05$, and an interaction between the three (sung and key-press direction, interval key, and spectral series): $F(6,438) = 5.53, MS_e = 0.03, p =$

$0.0001, \eta^2 = 0.01$. In general, sung accuracy was less influenced by spectral series and interval key than was key-press accuracy. As in the other analyses, participant accuracy worsened as the interval between the tones got smaller, and the discord between sung and key-press direction became evident as well. Again, musicians and non-musicians performed the most accurately in the Middle spectral series.

5.3.4 Sung Frequencies of the Reference Tones

The fundamental frequencies of the reference tone vocalized in the *vocal reproduction* condition for each participant were isolated using PRAAT (Boersma & Weenink, 2011). Figure 5.4 shows the sung frequency of the reference tone - tone F_0 (corrected for octave) in the vocal reproduction condition of Experiment 1 for both musicians and non-musicians. As shown, both musicians and non-musicians were fairly accurate and not statistically significantly different in their vocal reproductions of the reference tone (C4). This is not surprising, as C4 was the reference tone for every tone pair used in this experiment. There were no statistically significant differences be- tween the two participant groups. However, both groups were sharp in their vocal reproductions of the reference tone when the test tone occurred below middle C, and their vocal reproductions were flatter when test tones were above middle C, although this was not statistically significant.

Both musicians and non-musicians were influenced by spectral series, with the most accurate performance occurring in the Middle spectral series: $F(1,108) = 6.81, MS_e = 115.63$, $p = 0.0001, \eta^2 = 0.004$. Participants were also influenced by interval key, with performance accuracy decreasing as the intervals between the two tones became smaller: $F(2,186) = 11.69$, $MS_e = 469.82$, $p = 0.0001$, $\eta^2 = 0.04$.

5.3.5 Sung Frequencies of the Test Tones

The fundamental frequencies of the test tone vocalized in the *vocal reproduction* condition for each participant were isolated using PRAAT (Boersma & Weenink, 2011). However, the results of the sung frequencies were influenced by participant performance; that is, some participants did not vocally reproduce tones for all trials (skipping some tones) or the recording of their vocal reproductions were too faint or disrupted, resulting in trials that were unable to be analyzed.

Figure 5.5 shows the sung frequency of test tone - tone F_0 (corrected for octave) in the vocal reproduction condition of Experiment 1 for both musicians and non-musicians. As shown, both musicians and non-musicians were slightly less accurate in their vocal repro- ductions of the test tone, and there were no statistically significant differences between the two groups. Again, both musicians and non-musicians were influenced by spectral series, with the most accurate performance occurring in the Middle spectral series: $F(1,125) = 10.40$, $MS_e = 254.16$, $p = 0.0001$, $\eta^2 = 0.008$. Participants were also influenced by interval key, with performance accuracy decreasing as the intervals between the two tones became smaller: $F(3,214) = 27.65$,$MS_e = 1457.72$, $p = 0.0001$,$\eta^2 = 0.17$.

5.4 DISCUSSION

Experiment 1 presented tone pairs (constructed without the F_0) to musicians and non-musicians. The tones themselves were ramped in that they had a quiet onset, peak in volume, and ramped offset. The presentation of the tones were counterbalanced in that each participant started on one of the four conditions, attempting to prevent the instructions of any one condition influencing responses in the subsequent conditions.

When looking at the overall analyses of the four conditions, musicians were significantly more accurate than non-musicians, with both groups performing the most accurately with tones constructed via the Middle spectral series. Overall, participants performance was the most accurate in the *vocal reproduction* condition and least accurate in the *slider* condition, though not significantly. Contrary to previous research, vocally reproducing the tone pairs (and therefore introducing the respective F_0 of the tone pairs) did *not* significantly help participants identify the direction of tone pairs (Granzow, 2010; Reimchen, 2011).

Regardless of condition (and analyses), both musicians and non-musicians performed best in the Middle spectral series, when both the implied F_0 and the overtones were congruent. Both musicians and non-musicians were influenced by an interaction between spectral series and interval key. In particular, participant accuracy declined considerably on piano key number 38 with tones constructed through the High spectral series, and piano key number 42 with tones constructed through the Low spectral series. Participant performance accuracy declined around piano keys number 38 and number 42 in the Middle spectral series as well, though not significantly. This indicates that the congruent movement of the implied F_0 and overtones were required in order to successfully complete this task.

Both musicians and non-musicians portrayed a discord between sung direction and key-press. In particular, participants were more likely to sing the tones in the correct direction and indicate their key-press response in the wrong direction, especially as the intervals between the tones became smaller. Participants were influenced by spectral series, again having concordant sung and key-presses for tones constructed via the Middle spectral series.

Musicians and non-musicians did not differ as significantly as predicted, nor did their performance mirror results in the literature, with respect to differences in performance accuracy noted between musicians and non-musicians (regardless of response instruction) (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). Seither-Preisler et al. (2007), Granzow (2010), and Reimchen (2011) all had specific criteria in classifying participants as either professional musicians, amateur musicians, and non-musicians. In the two experiments conducted in this thesis, participants were classified as either a musician or non-musician based on their responses to an informal interview conducted by the researcher. Therefore, it is likely that some participants that would have been classified as “amateurs” in previous experiments may have been classified as either a musician or a non-musician in this thesis, thereby impacting the overall responses of each group.

Measuring musical expertise can be extremely subjective. An individual can learn to play guitar by ear, never knowing how to read music, and be considered a musician. In contrast, someone can complete formal musical training (including music theory, history, and so on) and not identify as a performer. As it currently stands, there is no short and accurate measurement for musical expertise and competence, as expertise can be extremely varied. It can also be difficult to compare musical expertise of different musical genres. For example, a rock musician may not have completed courses in music theory to the level of a classical violinist, and a classical musician is less likely to understand or engage in jazz charts and improvisation. In these experiments, participants self-identified as musicians and were dichotomized as musician and non-musician through the informal interview.

Both musician and non-musician accuracy dropped as the interval between the reference tone and test tone got smaller. As previously mentioned, pitch occurs on a spectrum (much like colour). As pitch frequencies get closer together, we more likely to want to perceptually classify them into the same group and it can create difficulty in identifying pitch direction.

There are a few reasons why the performance accuracy of musicians and non-musicians did not differ as significantly as in the literature. The most likely reason is that unlike in previous research, the discrimination between musicians and non-musicians was not as definitive. In previous research, participants were identified as either a professional musician, an amateur, or a non-musician. In this thesis, participants were only identified as musicians and non-musicians via an informal interview. Therefore it is likely that individuals who would have previously been identified as amateurs would have been identified as a musician or non-musician, thereby masking the performance of "true" professional musicians and "true" non-musicians.

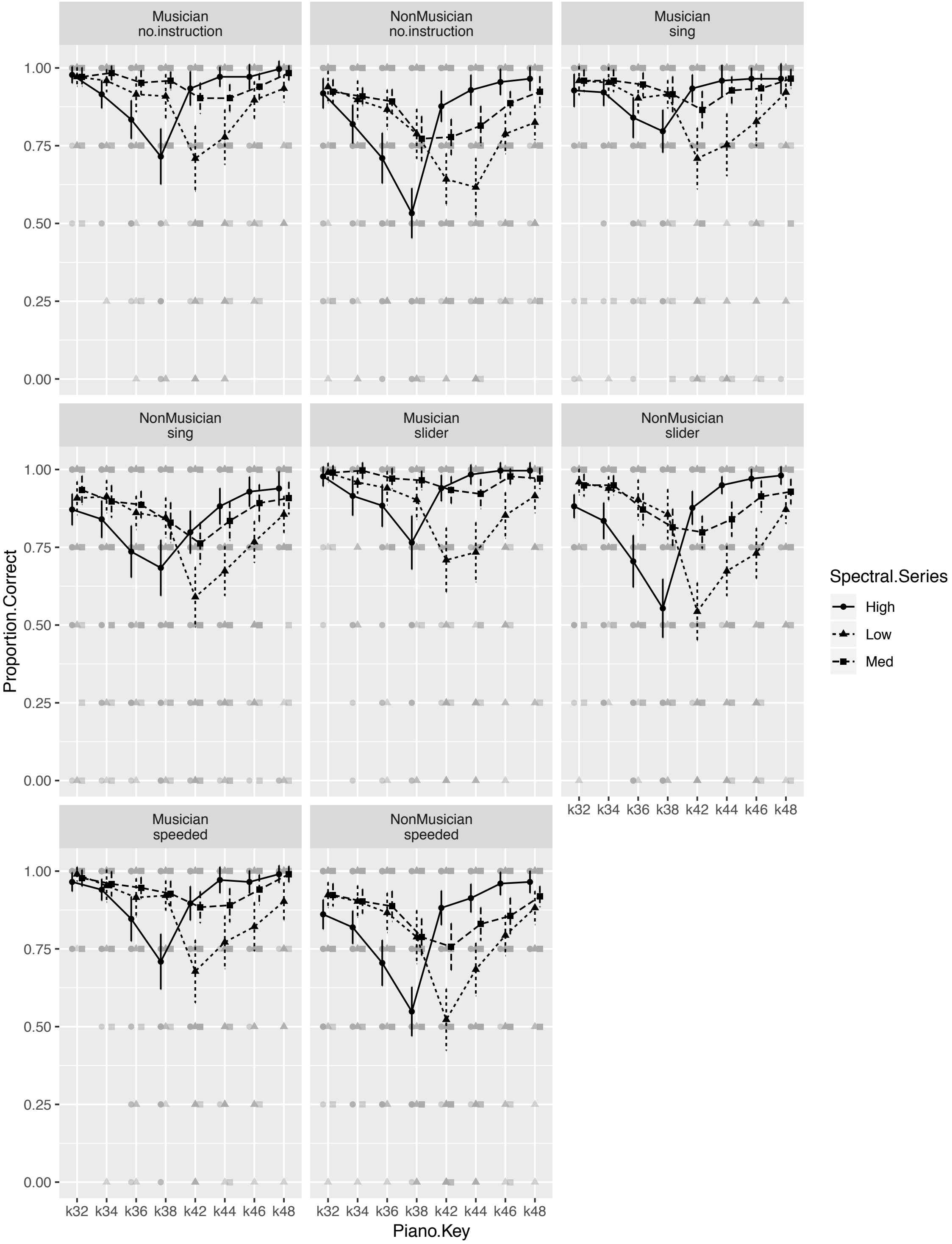

Figure 5.1: Mean proportion correct – concordant with F_0 – as a function of Test Condition, Spectral Series, and Piano Key number for both musicians and non-musicians in Experiment 1.

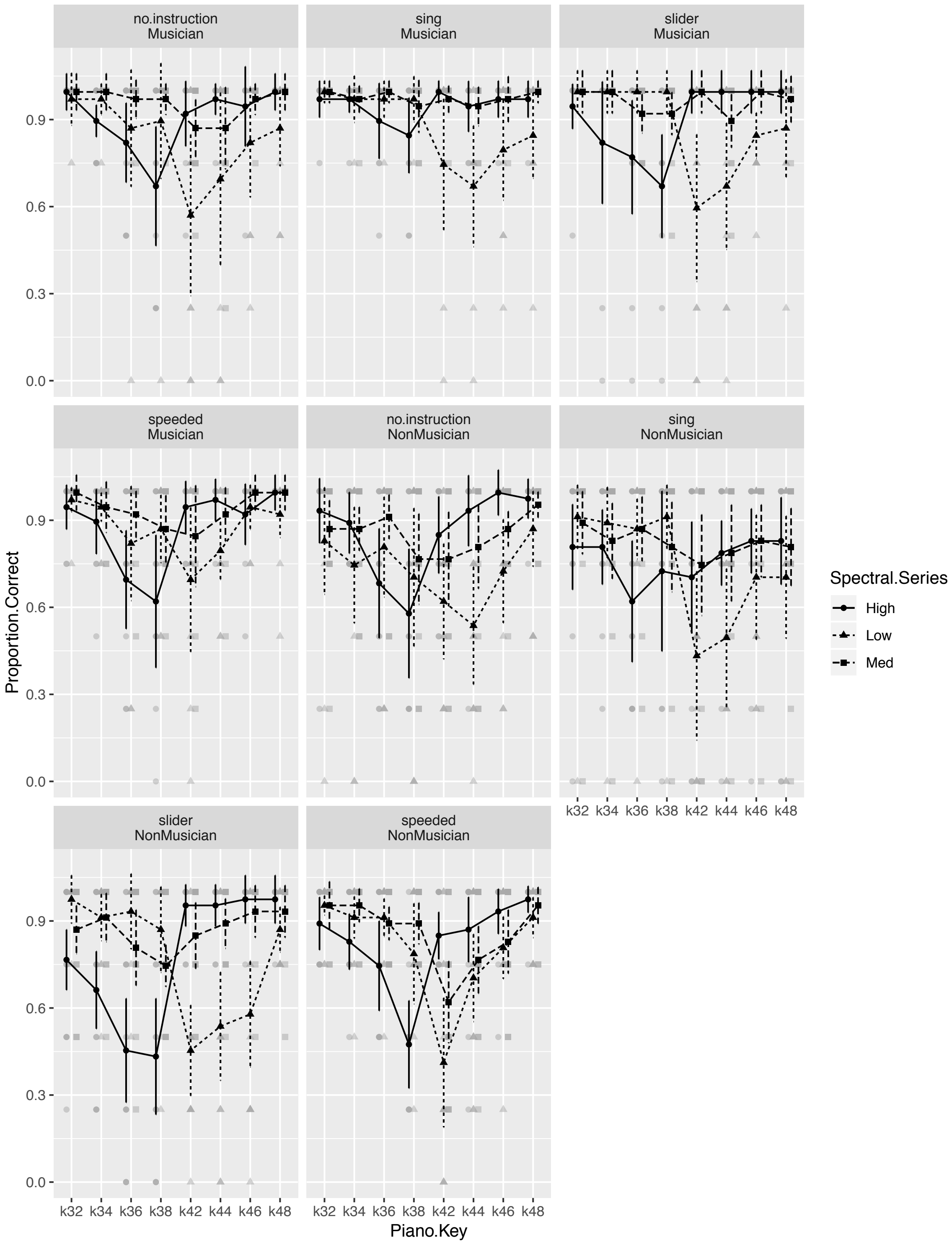

Figure 5.2: Mean proportion correct – concordant with F_0 – as a function of Counterbalancing Condition, Spectral Series, and Piano Key number for the initial Test condition of both musicians and non-musicians in Experiment 1.

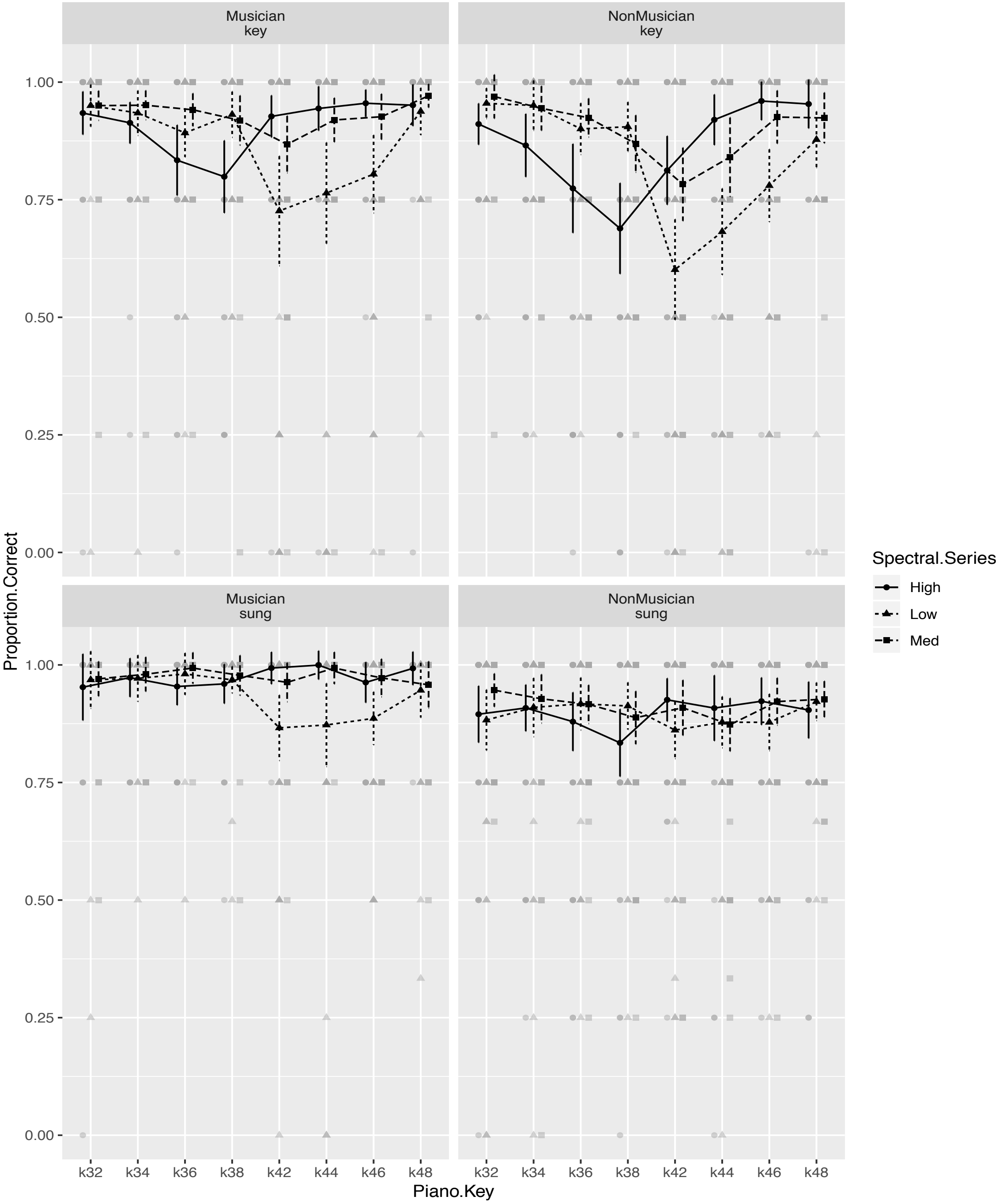

Figure 5.3: Key-Press versus Sung Direction in the Vocal Reproduction Condition of Experiment 1.

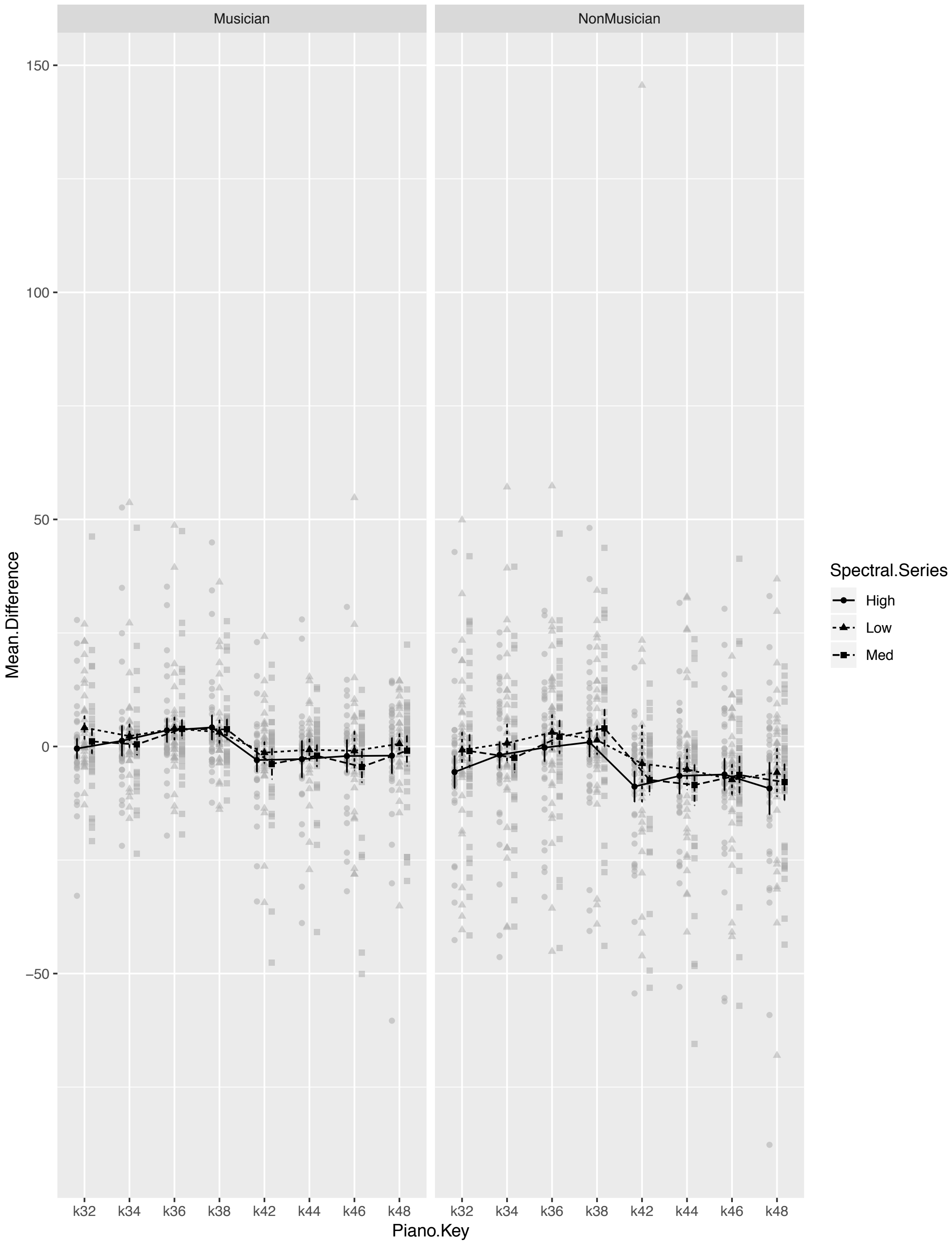

Figure 5.4: Sung Frequency of Reference Tone – Tone F_0 – (corrected for octave) in the vocal reproduction condition of Experiment 1.

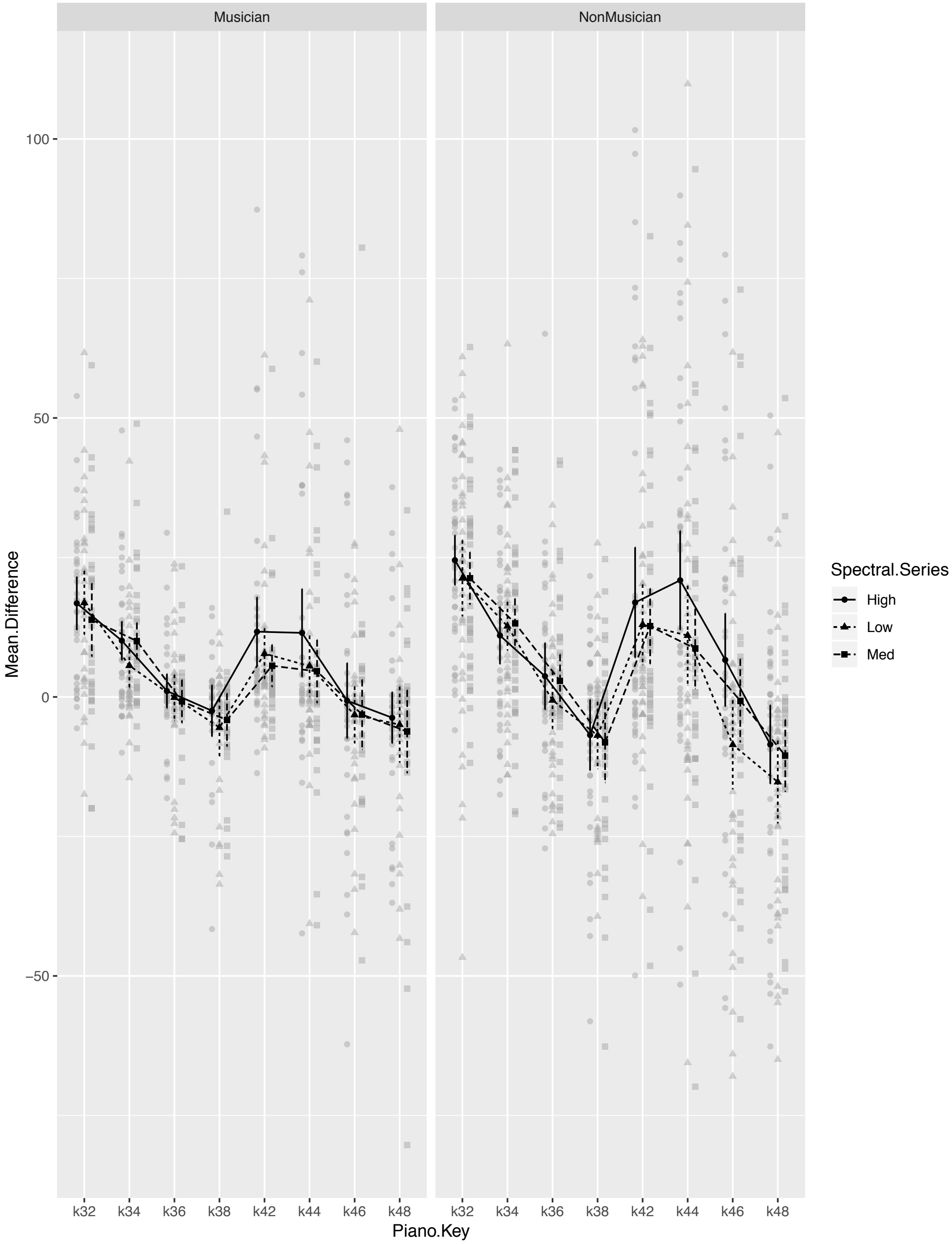

Figure 5.5: Sung Frequency of Test Tone – Tone F_0 (corrected for octave) in the vocal reproduction condition of Experiment 1.

CHAPTER SIX: EXPERIMENT 2: NON-RAMPED TONES

Results from Experiment 1 were not what we predicted, nor did they mimic results observed in the literature. Previously, it was strongly established that musicians consistently outperformed non-musicians, and that the inclusion of vocally reproducing the tones before indicating responses aided in performance accuracy (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). We hypothesized the results from Experiment 1 may have been due to the nature of the tones; specifically, the ramping of the tones. In Experiment 1, the tones were ramped in that they began quietly and slowly increased in volume until the tone "peaked" in intensity (also known as the apex). It was at this point that the full amount of auditory information was presented to participants. The tones then followed a ramped offset and subsequent withdrawal of the full amount of auditory information. In Experiment 1, the use of ramped tones may not have given the participant enough time to truly hear the tones. Therefore, there is a chance that performance was negatively influenced by the lack of full exposure received in regard to the stimuli. In order to account for this potential confound, we decided to run the same four conditions on musicians and non-musicians using tones that did not follow the Gaussian curve, while also taking order effects into account by analyzing the conditions as a between-subjects design, as opposed to a within-subjects design. The formation of the tones mirrored those used by Granzow (2010).

6.1 PARTICIPANTS

8 non-musicians and 13 musicians completed this experiment. Individuals were seated at a computer monitor and keyboard. They were given headphones as well as instructions on how to adjust the volume of the tones. After indicating their consent, they

were then instructed to begin the experiment. Vocal reproductions were recorded via microphone.

6.2 MATERIALS AND PROCEDURE

The tones outlined in chapter 4 were the stimuli in this experiment. The tones were digitized to 8-bit WAV files with a sampling frequency of 44,100 Hz, a duration of 500 msec, and a square- wavelength. The tones in this experiment did not have a quiet onset, peak, and tailed offset (as in Experiment 1). Instead, they began and ended abruptly, with full auditory information occurring for the full duration of the tone (as opposed to only milliseconds of full auditory information at the apex of the ramped tone). Abrupt offsets, however, allow for a "ringing" of the tone, even after the tone has finished playing. Previous research has shown that this "ringing" prolongs the duration of the auditory image, perhaps enabling participants to "receive" more auditory information than with ramped tones (Grassi & Darwin, 2006).

Unlike the counterbalanced design in Experiment 1, participants in Experiment 2 all began on condition 4 (the *no instruction* condition), as the ordering of conditions had no significant effect in the previous experiment. We were also attempting to re-establish the phenomenon found in Granzow (2010) and Reimchen (2011); particularly, the extreme differences in accuracy found between musicians and non-musicians. Having all the participants begin the study with the *no instruction* condition most closely mirrored Reimchen (2011).

6.3 RESULTS

The experiment resulted in two different measures of performance: (1) judgements of pitch direction as given by the key-press and slider responses, and (2) the actual frequencies produced by participants in the *vocal reproduction* condition, and how those vocal reproductions compared to key-press responses. For the key-press responses (the *no instruction* condition, the *speeded* condition, and the *vocal reproduction* condition), a response scored as concordant if it was consistent with the direction of the two tones; for the *vocal reproduction* condition, a key response was scored as concordant with "/" associating with the "up" direction and "z" associated with the "down" direction of the tones for that trial. These were then summed and converted to proportions of concordant responses for each experimental condition.

6.3.1 Proportion Correct

The proportion correct—concordant with F_0—was subjected to a 4 (Counterbalancing Group) × 2 (Musician Type: musician, non-musician) × 4 (Test Condition: no instruction, sing, slider, speeded) × 3 (Spectral Series: high, low, middle) × 8 (piano Key number: 32, 34, 36, 38, 42, 44, 46, 48) mixed analysis of variance with participants crossing Test Condition, Spectral Series, and Piano Key, but nested within Counterbalancing Group and Musician Type. The Greenhouse-Geisser degrees of freedom correction for sphericity was used for all effects involving within-participant factors.

Shown in Figure 6.1 are the mean proportion correct—concordant with F_0—as a function of test condition, spectral series, and piano key number for both musicians and non-musicians in Experiment 2. Overall, there were no statistically significant differences between musicians and non-musicians. However, participants were influenced by condition: $F(3,57) = 2,73, MS_e = 0.41, p = 0.0001, \eta^2 = 0.02$. Overall, participants performed the most accurately in the *vocal reproduction* condition. Participants were also influenced by spectral series (performing the most accurately in the Middle spectral series): $F(2,38) = 5.59, MS_e = 0.12, p = 0.0001$. There was also an influence of piano key number, with performance accuracy worsening as the intervals between keys became smaller: $F(7,133) = 9.30, MS_e = 0.08, p = 0.0001$. Participants were also influenced by an interaction between interaction between condition and spectral series: $F(6,114) = 2.04, MS_e = 0.03, p = 0.0001$. There was also an interaction between condition and interval: $F(21,399) = 1.47, MS_e = 0.03, p = 0.0001$. Participant performance was influenced by an interaction between spectral series and interval key: $F(14,266) = 5.49, MS_e = 0.07, p = 0.0001$. Lastly, performance accuracy was influenced by an interaction between condition, spectral series, and interval size (performing best in the *vocal reproduction* condition with tones presented at a larger interval constructed via the Middle spectral series): $F(72,798) = 1.58$, $MS_e = 0.02, p = 0.0001$.

6.4 VOCAL REPRODUCTION

Three analyses were conducted on the vocal reproductions of the participants: (1) how concordant was the vocal production of the tones with the direction of the tone pairs (similar to the key-press response analysis); (2) how accurately did participants reproduce the frequency of the reference tone (middle C, or C4), and (3) how accurately did they reproduce the test tone of each tone pair?

6.4.1 Key Press versus Sung Direction

The fundamental frequencies of the two tones (reference tone and test tone) vocalized in the *vocal reproduction* condition for each participant were isolated using PRAAT (Boersma & Weenink, 2011). Overall, there were no statistically significant differences between musicians and non- musicians, although both groups were influenced by spectral series ($F(1,32) = 2.97, MS_e = 0.05, p = 0.0001, \eta^2 = 0.02$) and interval key

($F(4,80) = 2.39, MS_e = 0.02, p = 0.0001, \eta^2 = 0.02$). Performance accuracy was highest with tones constructed through the Middle spectral series and performance accuracy worsened as the intervals between tones became smaller (particularly piano keys 38 and 42).

Overall performance accuracy for both musicians and non-musicians was influenced by spectral series, with participants performing more accurately with tones constructed through the Middle spectral series: $F(1,32) = 2.97, MS_e = 0.05, p = 0.0001, \eta^2 0.02$. Participants were also influenced by interval key, with performance accuracy worsening as the intervals between the tones became smaller: $F(4,80) = 2.39, MS_e = 0.02, p = 0.0001, \eta^2 = 0.02$.

6.4.2 Sung Frequencies of the Reference Tones

The fundamental frequencies of the reference tone vocalized in the *vocal reproduction* condition for each participant were isolated using PRAAT (Boersma & Weenink, 2011). Shown in Figure 6.4 are the mean sung frequencies (Hz.) of the reference tones - tone F_0 (corrected for octave). There were no statistically significant differences between musicians and non-musicians in their vocal reproductions of the reference tone. As seen in Experiment 1, both groups were sharp in their vocal reproductions of the reference tone when the test tone occurred below middle C, and their vocal reproductions were flatter when test tones were above middle C, although this was not statistically significant.

Participants were influenced by piano key, number with performance accuracy decreasing as the intervals between the two tones became smaller: $F(1,33) = 8.51, MS_e = 531.02, p = 0.0001, \eta^2 = 0.09$.

6.4.3 Sung Frequencies of the Test Tones

The fundamental frequencies of the test tone vocalized in the *vocal reproduction* condition for each participant were isolated using PRAAT (Boersma & Weenink, 2011). Shown in Figure 6.5 are the mean sung frequencies (Hz.) of the test tones - tone F_0 (corrected for octave). Both musicians and non-musicians were slightly less accurate in their vocal reproductions of the test tone when compared to the vocal reproductions of the reference tone, and there were no statistically significant differences between the two groups of participants.

Again, both musicians and non-musicians were influenced by spectral series, with the most accurate performance occurring in the Middle spectral series: $F(1,30) = 4.38, MS_e = 312.82, p = 0.0001, \eta^2 = 0.01$. Participants were also influenced by interval key, with performance accuracy decreasing as the intervals between the two tones became smaller: $F(3, 62) = 10.97, MS_e = 1431.83, p = 0.0001, \eta^2 = 0.23$.

6.5 DISCUSSION

As the results of Experiment 1 did not mirror results found in the literature, we included Experiment 2 as a means to attempt to replicate previous findings. In particular, we wanted to see if we could influence musicians to perform with higher accuracy and if we could manipulate the performance of non-musicians to fall more to chance, while simultaneously observing aspects of vocal reproduction with the addition of the two new paradigms (the *slider* condition assessing the possible embodiment of pitch perception and the *speeded* condition, assessing performance accuracy when all aspects of vocal reproduction were removed) (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). The ramping of the tones was changed (instead presenting them without any ramping) in an effort to give participants the full amount of auditory information for the duration of the tone, and to mirror the stimuli used by Granzow (2010).

Experiment 2 was identical to Experiment 1, except for the ramping of the tones and presentation of conditions. Experiment 2 used tones constructed with an abrupt onset and offset with no change in volume (thereby presenting full auditory information for the entire duration of the tone). Instead of counterbalancing the conditions among participants (as in Experiment 1), both musicians and non-musicians began on the *no instruction* condition, as this was the most common paradigm used in the literature, followed by the *vocal reproduction* paradigm (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). As all participants began on the *no instruction* condition, only the proportion correct (pc) of performance accuracy of participants across all four conditions (as a within-subjects design) was analyzed.

As in Experiment 1, there were no significant differences between musicians and non- musicians. Even with the non-ramped stimuli, musicians and non-musicians still performed relatively well. Performance accuracy of both musicians and non-musicians was highest in the Middle Spectral series. There were no differences between musicians and non-musicians in regard to sung direction, although participants were still influenced by spectral series (with both groups performing the most accurate in the Middle Spectral series) and interval key (performance accuracy decreased as the interval between the tones became smaller, particularly around keys 38 and 42).

Unlike Experiment 1, there were no differences between musicians and non-musicians in key-press versus sung direction performance. In Experiment 1, musicians were more accurate in their overall vocal reproductions, and performance of both musicians and non-musicians were influenced by interval key and spectral series. In

Experiment 2, participant performance was not influenced by participant type (musician and non-musician) but was influenced by spectral series and interval key.

In regard to the actual sung key of the reference tone, there were still no statistically significant differences between musician and non-musician accuracy, although participants were influenced by interval key. Accuracy of the sung reference note seemed to be influenced by the test tone, particularly if the test tone was close (in interval) to the reference tone. Participants were presented with the two tones and then asked to vocally reproduce them; therefore, it is likely that the vocal reproduction of the reference tone was influenced by the test tone. There were also no statistically significant differences between musicians and non-musicians in the sung accuracy of the test tone. As with the reference tone, participants were influenced by interval key as well as spectral series.

As in Experiment 1, the discrimination abilities between musicians and non-musicians was not as distinct as in the literature (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). Participants self-identified as musicians and were dichotomized as musician and non-musician through the informal interview, as opposed to stricter criteria and the addition of a third group of participants (musical amateurs). In Experiment 1, there were more significant differences between musicians and non-musicians, particularly in the accuracy of the vocal reproductions. In Experiment 2, there were no statistically significant differences between musicians and non-musicians, although both musicians and non-musicians were still influenced by spectral series and interval key (as in Experiment 1).

As in Experiment 1, both musicians and non-musicians were influenced by spectral series, regardless of condition. Both musicians and non-musicians performed best when tones were constructed through the Middle spectral series, indicating that the congruent movement of the implied F_0 and overtones were required in order to successfully complete this task. The worst performance occurred on lower pitch intervals (with test tones falling below piano key number 40) constructed through the High spectral series. In regard to the Low spectral series, performance was worse on higher pitched intervals (intervals above piano key number 40). Participants (both musicians and non-musicians) were also influenced by interval key, with performance accuracy worsening around piano keys 38 and 42.

This thesis differed from previous experiments in a variety of ways. In Experiment 1, the tones generally mirrored the gaussian formation used by Reimchen (2011) and Seither-Preisler et al. (2007). In Experiment 2, the tones mirrored those of Granzow (2010), with an abrupt onset, sustained tone, and abrupt offset. Unlike the previous experiments, participants were dichotomized as musicians or non-musicians via an informal verbal interview (as opposed to a strict survey). In Experiment 1, the conditions presented to participants were counterbalanced so that the instructions and performance in one condition did not influence the performance in another. In particular, we hypothesized that once participants completed the *vocal reproduction* condition, they would use this "trick" to aid in performance accuracy in the remaining conditions, regardless of instruction, as Granzow (2010) and Reimchen (2011) noted an immediate increase in performance accuracy once participants were instructed to vocally reproduce the tones.

For both experiments, the tones were digitized to 8-bit WAV files with a sampling frequency of 44,100 Hz, with a duration of 500 msec and constructed without the F_0. In

Experiment 1, the onset of the tones were ramped, in that they slowly reached peak volume (and auditory information) at the apex (peak) of the ramp followed by a ramped offset. In Experiment 2, tones began and ended abruptly, with full auditory information occurring for the full duration of the tone (as opposed to only milliseconds of full auditory information at the apex of the ramped tone). In Experiment 1, musician performance was (overall) more accurate than non-musicians, and in Experiment 2, there were no differences between musician and non-musician performance. The differences in performance accuracy between the two experiments were likely *not* due to the ramped (or non) nature of the tones. In Experiment 2, the musicians performed slightly worse than in Experiment 1, and non-musicians were slightly more accurate than in Experiment 1. Therefore, the difference in results between the two experiments is likely not due to a difference in performance accuracy but a skewed grouping of “musician” and “non-musician,” and the likely inclusion of “amateur” musicians in each of the two groups.

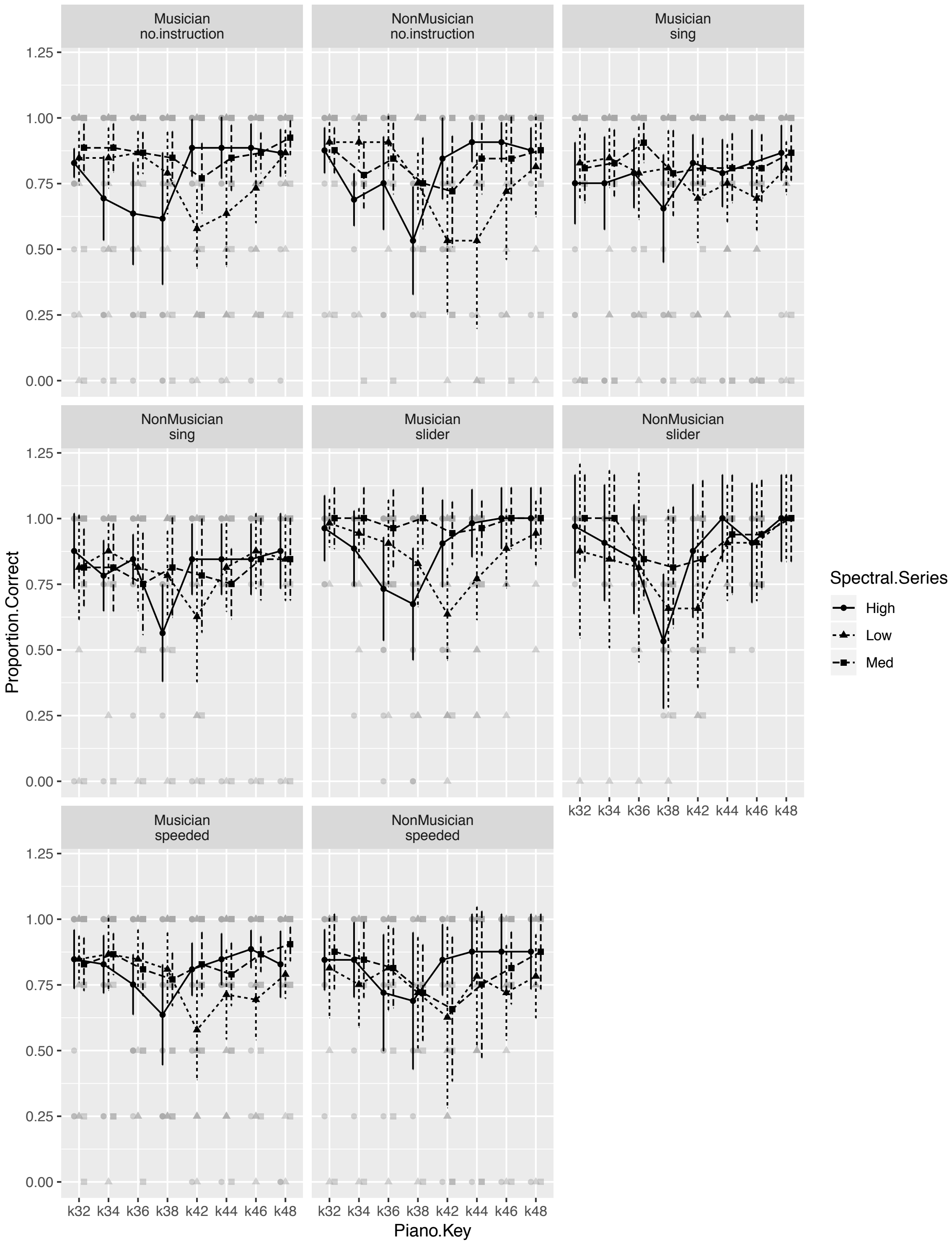

Figure 6.1: Mean proportion correct – concordant with F_0 – as a function of Test Condition, Spectral Series, and Piano Key number for both musicians and non-musicians in Experiment 2.

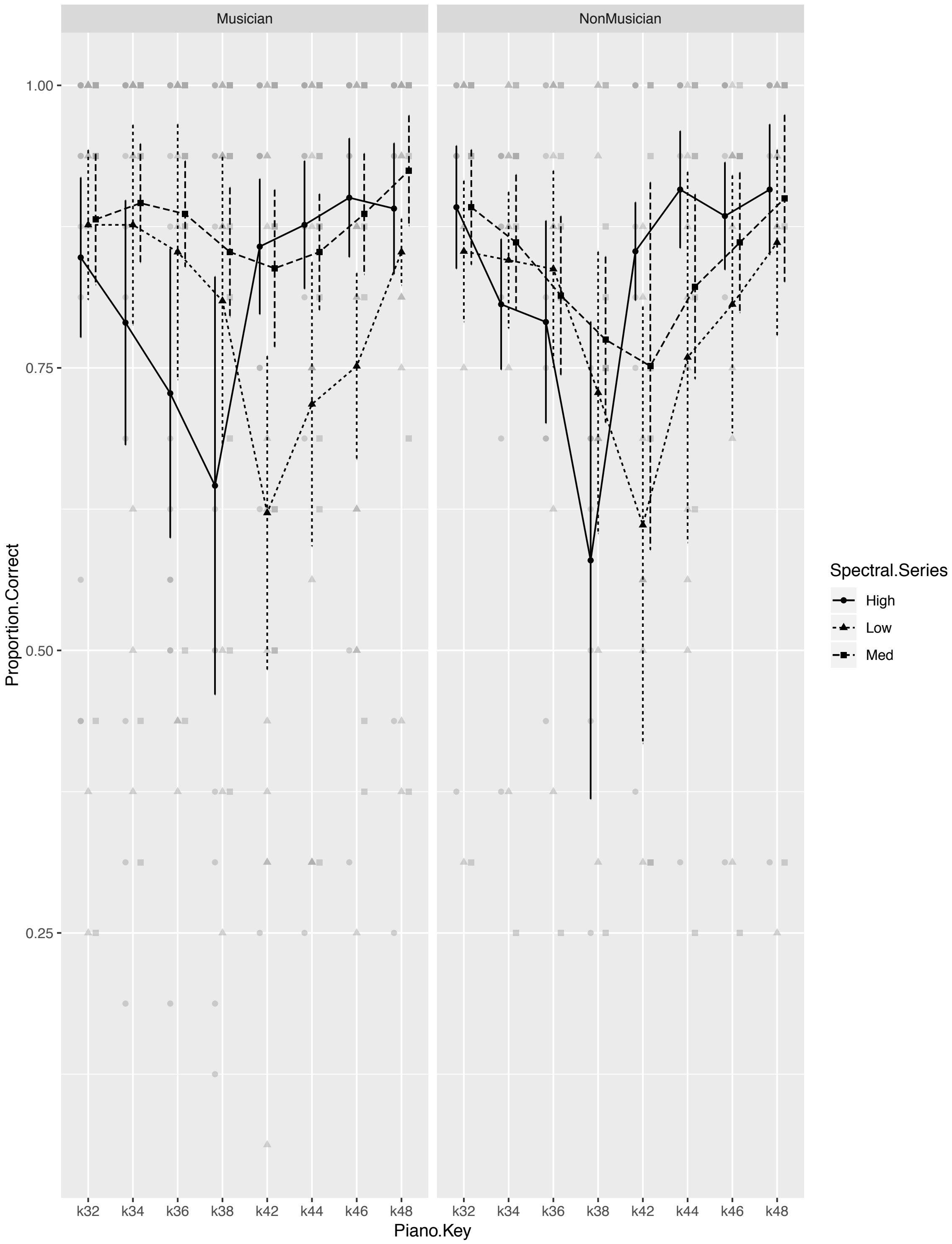

Figure 6.2: Mean proportion correct Sung Direction in the Vocal Reproduction Condition for musicians and non-musicians in Experiment 2.

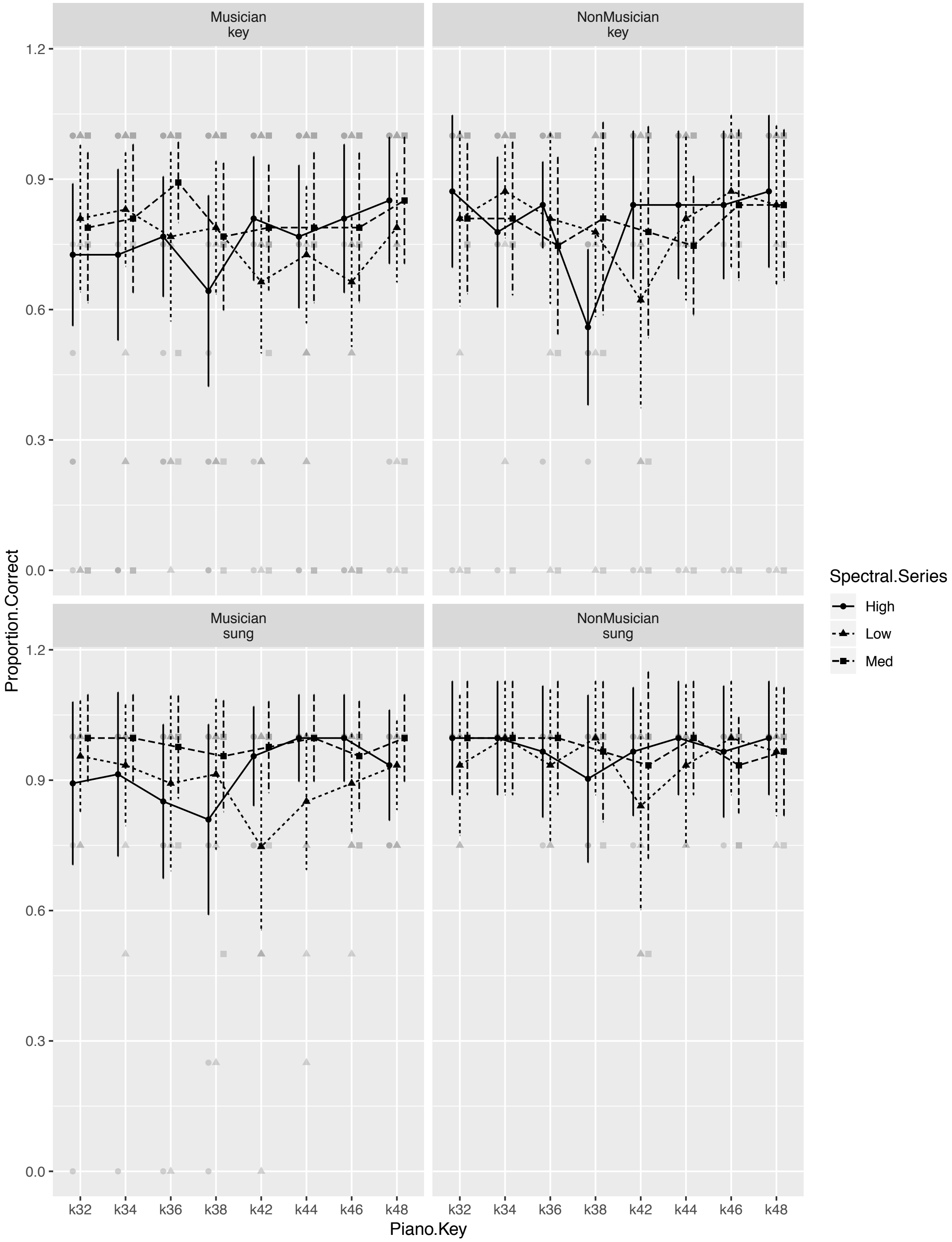

Figure 6.3: Mean proportion correct Key-Press versus Sung Direction in the vocal reproduction condition for musicians and non-musicians in Experiment 2.

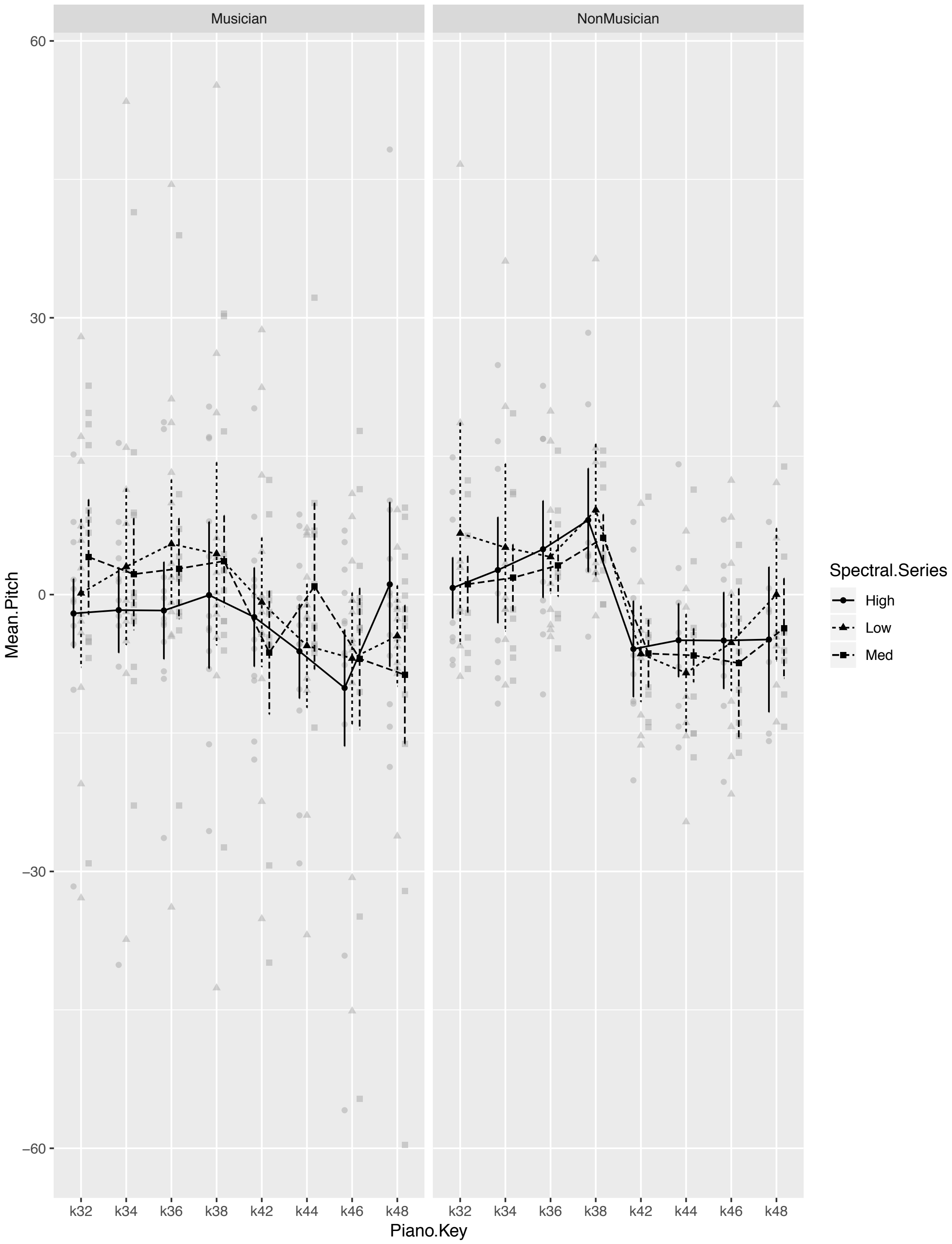

Figure 6.4: Mean Sung Frequency (Hz.) of Reference Tone – Tone F_0 (corrected for octave)

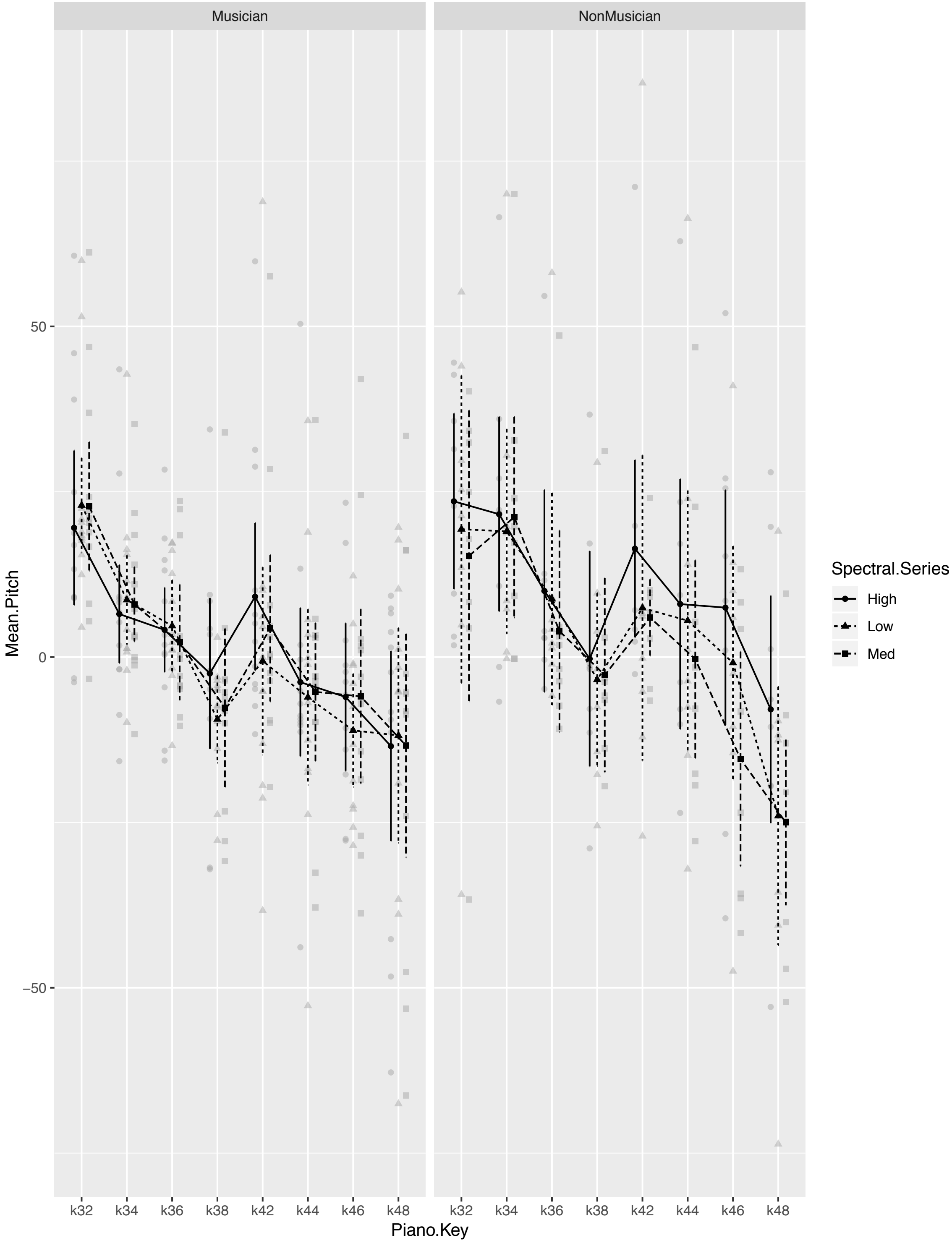

Figure 6.5: Mean Sung Frequency (Hz.) of Test Tone – Tone F_0 (corrected for octave)

CHAPTER SEVEN: CONCLUSION

In experiments on pitch discrimination completed prior to this thesis, the common paradigm was to present participants with tone pairs and ask them to identify whether the test tone was higher or lower than the reference tone via computer key-presses, with one key indicating the test tone went up and another key indicating it went down (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). Previous research also examined the influence of vocal reproduction of the tones prior to responding with computer key-presses in performance accuracy (Granzow, 2010; Reimchen, 2011). Previous studies established that requesting participants (both musicians and non-musicians) to vocally reproduce tones in a pitch perception task (using tones constructed without the F_0) resulted in higher performance accuracy (Reimchen, 2011). Accordingly, we expected to mirror the results of the previous studies (particularly, Granzow (2010); Reimchen (2011)).

By including two new paradigms (*slider* and *speeded*), we were able to measure whether or not different responses to a pitch discrimination task aided in accuracy and the ability to identify *what* aspects of vocal reproduction aided in performance accuracy; is the reintroduction of the F_0 in addition to the muscular response necessary in order to successfully complete this task (assessed via the *vocal reproduction* condition), or is it sufficient to attempt to mirror subtle muscular response without the actual vocal reproduction of the tone (as in the *slider* condition)? Or, perhaps, performing this task in a speeded manner (thereby preventing any aspect of vocal reproduction from aiding in performance accuracy) allowed participants to successfully complete this task (as determined by the *speeded* condition), as previous research has established a link between speed and accuracy in pitch discrimination tasks (Mulder et al., 2013; Duckworth, Potticary, & Badyaev, 2018; Gold & Shadlen, 2007).

The *slider* condition was included as a means to potentially mimic laryngeal movement that occurs when an individual sings, without having the participant vocally reproduce the tones (and subsequently reintroduce the F_0). The *speeded* condition was included as a means to suppress the effects of vocal reproduction (the reintroduction of the F_0 as well as muscular (laryngeal) feedback that occurs with singing). By asking participants to respond as quickly as possible, we hypothesized that we would be able to truly narrow down whether aspects of vocal reproduction (the reintroduction of the F_0 and laryngeal movement) were required to accurately complete this task. If the reintroduction of the F_0 helped participants perform accurate judgments, then they would have performed with the most accuracy in the *vocal reproduction* condition, mirroring the performance cited by Reimchen (2011). If, instead, they were able to transfer muscular (laryngeal) feedback into perceptual judgments and attend to their vocal reproductions, performance would have been the most accurate in the *vocal reproduction* condition. If performance was the most accurate in the *slider* condition, it could be due to the possibility that participants were able to attend to the muscular feedback and project that feedback onto a computer screen. If performance was best in the *speeded* condition, then it could be concluded that neither aspect of vocal reproduction (the reintroduction of the F_0 and laryngeal movement) aided in performance accuracy, which would be contrary to previous research.

The stimuli used in this thesis mirrored the spectral construction of previous research (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). Seither-Preisler et al. (2007) developed the Auditory Ambiguity Test (AAT), which included two spectral series in which the implied F_0's and overtones moved in opposition; the High and Low spectral series. Granzow (2010), mirroring the experiment conducted by Seither-Preisler et al. (2007), included a third spectral series, the Middle spectral series, in which the implied F_0's and overtones always moved together (in concert). The Middle spectral series was included in order to allow for the assessment of which aspects of pitch participants were responding to; the implied F_0's or the overtones (Granzow, 2010). Granzow (2010) noted an increase in performance accuracy when non-musicians were asked to vocally reproduce the tones, although he did not record the vocal reproductions. Reimchen (2011) used the three spectral series outlined in Granzow (2010), but instead altered the responses participants were required to make by including a paradigm requiring vocal reproduction of the tones and recording the reproductions. These reproductions, as in this thesis, were analyzed in conjunction with the computer key-presses to assess discrimination capabilities among the participants.

In this thesis, it was predicted that both musicians and non-musicians would have the most trouble in the Low spectral series, followed by the High series, as the manipulation (having the overtones and implied F_0's move in opposition) could produce conflicting auditory stimuli. In contrast, we expected participants to perform most accurately with tones constructed in the Middle spectral series, as natural sounds are constructed with overtones and F_0 moving in concert, without any conflict in the movement of the F_0 and associated overtones. In contrast, if participants were tracking overtone frequencies, then their accuracy would showcase patterns of response accuracy (in regard to semitone differences) in both the Low and High spectral series; in particular, we expected individuals to perform worse in lower pitched intervals (intervals occurring below C4) with tones constructed through the High spectral series, and in higher pitched intervals (intervals occurring above C4) constructed through the Low spectral series. We also predicted that performance accuracy would decrease as the intervals between the tones became smaller.

7.1 OVERALL CONCLUSIONS

When only attending to the proportion correct responses in Experiments 1 and 2, there are no statistically significant differences between the four conditions. In Experiment 1, musicians are significantly more accurate than non-musicians, in performance accuracy and vocal reproduction. In Experiment 2, however, there are no statistically significant differences between musicians and non-musicians. In both experiments, both musicians and non-musicians were influenced by overtone key and spectral series, as predicted. Performance accuracy was highest in the Middle spectral series throughout all four conditions, regardless of analyses or ramping of tones, indicating that the congruent movement of the implied F_0 and overtones were required in order to successfully complete a pitch discrimination task with tones constructed without the F_0.

Experiment 2 did not replicate the results from the overall analyses of Experiment 1 (in regard to musicians performing more accurately than non-musicians), suggesting that the order of the presentation of conditions may have played a role in response accuracy. Also, the previous studies used a between-subject design; the overall analyses of Experiment 1 was a within-subject design with the conditions being counterbalanced. It may be possible that the unlikely results of this analysis from Experiment 1 were due to the design itself (specifically, not taking order effects into account). Granzow (2010) also followed a within-subjects design in his analyses of Experiment 3 (where participants were asked to vocally reproduce the tones of *all* tone pairs), but as participants were not required to change how they responded (based on condition), this likely would not have had much of an effect on the results. Unlike (Granzow, 2010), the conditions included in Experiment 1 in this thesis all had the possibility of influencing responses in other conditions; in particular, participants may have found that vocally reproducing the tones (such as in the *vocal reproduction* condition) may have made the task easier, thereby applying this "trick" to other conditions (regardless of instruction).

It is likely that these results (of both Experiment 1 and 2) are due to the overlap of musical ability in participant groups. In previous research, participants were identified as either professional musicians, amateur musicians, or non-musicians (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). In this thesis, participants were only identified as musicians or non-musicians. It is likely that participants that would have previously been identified as amateur musicians would have been classified as either a musician or non-musician in this thesis, thereby affecting the overall performance of each participant group. Previous research defined professional musicians as individuals who have received over five years of musical training and were still regularly performing (Granzow, 2010). Seither-Preisler et al. (2007) identified musicians as individuals who had received education through a music conservatory and who engaged in regular practice. It is likely that in this thesis, an individual who has *not* received formal music education but *has* received over five years of musical training (perhaps via private lessons for a specific instrument) would be identified as a musician by Granzow (2010), but an amateur by Seither-Preisler et al. (2007). Therefore, by having participants classify themselves as musicians or non-musicians, it is likely that some participants who were classified as musicians in this thesis would have been classified as amateurs in previous research.

Musical expertise is extremely subjective; previous research has classified musical expertise in a variety of ways, including the completion of board-qualifying exams (for example, the Royal Conservatory of Music), the amount of time (years) spent engaging in musical practice, and receiving private instruction (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). However, these types of questions may not truly tap into true musicianship and expertise. There are several examples of professional musicians never receiving formal music training or even knowing how to read music (for example, Elvis Presley) (Patil, 2018). Therefore, by asking participants to dichotomize themselves as musicians or non-musicians, we would be able to account for *all* aspects of musical expertise.

7.2 LIMITATIONS AND FUTURE DIRECTIONS

The pitch discrimination accuracy of musicians and non-musicians did not differ to the degree expected in Experiment 1 and failed to reach significant at all in Experiment 2. Thus, we need to be concerned about failing to replicate the strong effects in the literature (Seither-Preisler et al., 2007; Granzow, 2010; Reimchen, 2011). Measuring musical expertise can be extremely subjective. An individual can learn to play guitar by ear, never knowing how to read music, and be considered a musician. In contrast, someone can complete formal musical training (including music theory, history, and so on) and not identify as a performer. As it currently stands, there is no quick and accurate measurement for musical expertise and competence, as expertise can be extremely varied, especially among different musical genres (e.g., a rock musician compared to a classical musician).

In these experiments, participants self-identified as musicians. In the future, it would be interesting to explore the role of more clearly defined and verified musical expertise and competence in relation to performance accuracy in a pitch discrimination with tones constructed without the F_0. In this thesis, most of the participants who identified as musicians were undergraduates recruited from the Department of Music at the University of Lethbridge. It would be interesting, once musical expertise and competence could be determined, to identify any possible differences between “amateur” musicians (such as university students) and “professional” musicians (such as full-time performers, recording artists, etc.). It would also be interesting to explore possible differences among different types of musicians. Nikjeh, Lister, and Frisch (2008) compared pitch discrimination abilities among non-musicians, instrumentalists, and vocalists. The performance of both groups of musicians was more accurate than the performance of the non-musicians; however, the vocalists were even more accurate in their responses when compared to the instrumentalists. It would be interesting to see how the two groups compare in tones lacking the F_0, or even compare specific instrumentalists (for example, a violin player compared to a saxophonist in their ability to discriminate between pitches lacking the F_0).

Both musician and non-musician accuracy dropped as the interval between the reference tone and test tone got smaller. As previously mentioned, pitch occurs on a spectrum (much like colour). It would be interesting to see if participants could identify intervals smaller than the ones used in these experiments (a minor second), and if their performance accuracy could be influenced by training. Since musicians are familiar with the Western categorization of semitones, making the tones closer together may inhibit their performance and force them to place the pitches in a category to which they believe it may not fully belong, perhaps altering their perception of the intervals and pushing the cultural boundaries of pitch perception.

REFERENCES

Aldrich, J. (1997). R.A. Fisher and the making of a maximum likelihood 1912-1922. *Statistical Science*, *12*, 162-176.

Bachem, A. (1955). Absolute pitch. *The Journal of the Acoustical Society of America*, *27*, 1180-1185.

Balaguer-Ballester, E., R. Clark, N., Coath, M., Krumbholz, K., & L. Denham, S. (2009). Understanding pitch perception as a hierarchical process with top-down modulation. *PLOS Computational Biology*, *5*, e1000301.

Ben-Haim, M. S., Eitan, Z., & Chajut, E. (2014). Pitch memory and exposure effects. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 24-32.

Bishop, J., & Keating, P. (2012). Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex. *The Journal of the Acoustical Society of America*, *132-2*, 1100-1112.

Boersma, P., & Weenink, D. (2011). *Praat: doing phonetics by computer [computer program]. version 6.0.24.* Retrieved from http://www.praat.org/

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

Brown, S. (2000). *The origins of music* (N. L. Wallin, B. Merker, & S. Brown, Eds.). The MIT Press.

Centres, D. H. A. (n.d.). *How we perceive sound.* Retrieved 2018, from https:// davidsonhearingaids.com/how-we-perceive-sound/

Cherry, K. (n.d.). *Gestalt laws of perceptual organization.* Retrieved November 2018, from https://www.verywellmind.com/gestalt-laws-of-perceptual-organization-2795835

Chialvo, D. R. (2003). How we hear what is not there: A neural mechanism for the missing fundamental illusion. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *13*, 1226.

Classroom, T. P. (2018). *Fundamental frequency and harmonics.* Retrieved 2018, from https://www.physicsclassroom.com/class/sound/Lesson-4/ Fundamental-Frequency-and-Harmonics

Collins. (2018). *phoneme.* Retrieved September 2018, from https://www.collinsdictionary.com/dictionary/english/phoneme

Coren, S., & Girgus, J. S. (1980). Principals of perceptual organization and spatial distortion: The gestalt illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 404-412.

Creighton, J. (2014). *Overtone singing: The science behind singing multiple notes at once.*

Cutnell, J. D., & Johnson, K. W. (1998). *Physics* (4th ed.). New York: Wiley.

Cyx, J., & Shapiro, M. (1986). Perception of missing fundamental by a species of songbird (sturnus vulgaris). *Journal of Comparative Psychology*, *100*, 356-360.

Deutsch, D., & Hawthorne, T. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception*, *21*, 339-356.

Duckworth, R. A., Potticary, A. L., & Badyaev, A. V. (2018). On the origins of adaptive behavioural complexity: Developmental channeling of structural trade-offs. *Advances in the Study of Behaviour*, *50*, 1-36.

Fourier, J.-B.-J. (1822). *Théorie analytique de la chaleur*. Paris: F. Didot.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535-574.

Gorow, R. (2011). *Hearing and writing music: Professional training for today's musician*. September Publishing.

Granzow, J. (2010). *Ventriloquial dummy tones: Embodied cognition of pitch direction* (Unpublished master's thesis). University of Lethbridge.

Grassi, M., & Darwin, C. J. (2006). The subjective duration of ramped and damped sounds. *Perception and Psychophysics*, *68*, 1382-1392.

Griffiths, T., Buchel, C., Frackowiak, R., & Patterson, R. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, *1*, 422-427.

Grody, A. (n.d.). *Why pitch matching is hard.*

Hall, D. A., & Palack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, *19*, 576-585.

Hamilton, J. (2009, October). *Say what?! musicians hear better.* Retrieved August 2018, from https://www.npr.org/templates/story/story.php?storyId=113938566

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, *11*, 465-472.

Harnard, S. (2003). *Categorical perception*. Cambridge University Press.

Heffner, H., & Whitfield, I. C. (1998). Perception of the missing fundamental by cats. *The Journal of the Acoustical Society of America*, *59*, 915-919.

Hoeschele, M. (2017). Animal pitch perception: melodies and harmonies. *Comparative cognition and behaviour reviews*, *12*, 5-18.

Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and non-singers. *Experimental Brain Research*, *190*, 279-287.

Klofstad, C. A., Nowicki, S., & Anderson, R. C. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 2698-2704.

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*, 578-584.

Kohler, K. (2012). The perception of lexical stress in German: Effects of segmental duration and vowel quality in different prosodic patterns. *Phonetica*, *69*, 68-93.

Krumbholz, K., Patterson, R., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, *13*, 765-772.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

Latinus, M., & Taylor, M. (2012). Discriminating male and female voices: Differentiating pitch and gender. *Brain Topography*, *25*, 194-204.

Lau, B., & Werner, L. (2012). Perception of the missing fundamental pitch by 3- and 4-month- old human infants. *The Journal of the Acoustical Society of America*, *132*, 3874-3882.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics*, *56*, 414-423.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358-368.

Lockhead, G., & Byrd, R. (1981). Practically perfect pitch. *The Journal of the Acoustical Society of America*, *70*, 387-389.

Loui, P., & Wessel, D. (2008). Learning and liking an artificial musical system: Effects of set size and repeated exposure. *Musicae Scientiae*, *12*, 207-230.

Lui, F., Chan, A. H. D., Ciocca, V., Roquet, C., Peretz, I., & Wong, P. C. M. (2016). Pitch perception and production in congenital amusia; evidence from Cantonese speakers. *The Journal of the Acoustical Society of America*, *140*, 563-575.

Ly, H. (2012). *Learning chinese: The importance of tones.*

Makin, S. (2014, March). *The genes that mean you have a good ear for music.* Retrieved August 2018, from https://www.newscientist.com/article/dn25206-the-genes -that-mean-you-have-a-good-ear-for-music/

Matsuwaki, Y., Nakajima, T., Ookushi, T., Iimura, J., Kunou, K., Nakagawa, M., . . . Ishikawa, T. (2004). Evaluation of missing fundamental phenomenon in the human auditory cortex. *Auris Nasus Larynx*, *31*, 201-211.

Missing fundamental effect. (n.d.). Retrieved from http://hyperphysics.phy-astr.gsu .edu/hbase/Sound/subton.html\#c2

Mulder, M. J., Keuken, M. C., van Maanen, L., Boekel, W., Forstmann, B. U., & Wagenmakers, E.-J. (2013). The speed and accuracy of perceptual decisions in a random-tone pitch task. *Attention, Perception, and Psychophysics*, *75*, 1048-1058.

Ngo, M., Vu, K., & Strybel, T. (2016). Effects of music and tonal language experience on relative pitch performance. *American Journal of Psychology*, *129*, 125-134.

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2008). Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. *Psychophysiology*, *25*, 994-1007.

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *Journal of the Acoustical Society of America*, *125*, 328-338.

Nooteboom, S. (1997). The prosody of speech: Melody and rhythm. *The Handbook of Phonetic Sciences*, *5*, 1-48.

O'Connor, K. (2019). *Tone deafness (amusia) and other causes of persistent pitch problems.*

Ogden, C. E., & Richards, I. A. (1923). *The meaning of meaning*. London: Kegan Paul.

Oxenham, A. J. (2012). Pitch perception. *The Journal of Neuroscience*, *32*, 13335-13338. Palisade. (2019). *Monte carlo simulation.* Retrieved from https://www.palisade.com/ risk/monte\carlo\simulation.asp

Patil, R. (2018). *10 legendary musicians who never learned how to read music.*

Retrieved March 26, 2018, from https://melodyful.com/legendary-musicians-who-never -learned-how-to-read-music

Patterson, R., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., & Allerhand, M. (1992). *Auditory physiology and perception* (Y. Cazals, L. Demany, & K. Horner, Eds.). Pergamon Press; Oxford, England.

Penagos, H., Melcher, J., & Oxenham, A. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, *24*, 6810-6815.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, *33*, 184-191.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness". *Music Perception: An Interdisciplinary Journal*, *25*, 95-115.

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, and Psychophysics*, *71*, 1385-1389.

Pinch, T., & Bijsterveld, K. (2012). *The oxford handbook of sound studies*. Oxford University Press.

Plack, C. J., Oxenham, A. J., & Fay, R. R. (2005). *Pitch: neural coding and perception*. Springer.

Popeil, L. (2015, December). *5 essential facts about your vocal range.* Retrieved December 2018, from http://www.voicecouncil.com/5-essential-facts-about-your-vocal-range/

Powers, H. S. (2003). *The harvard dictionary of music* (D. M. Randel, Ed.). Cambridge: Belknap Press for Harvard University Press.

Pujol, R., Nouvian, R., & Lenoir, M. (n.d.). *Hair cells: Overview.* Retrieved 19/09/2016, from http://www.cochlea.eu/en/hair-cells

Pure tones. (n.d.). Retrieved from http://www.indiana.edu/$\sim$p1013447/ dictionary/tone.htm

Qin, L., Sakai, M., Chimoto, S., & Sato, Y. (2005). Interaction of excitatory and inhibitory frequency-receptive fields in determining fundamental frequency sensitivity of primary auditory cortex neurons in awake cats. *Cerebral Cortex*, *15*, 1371-1383.

Reimchen, M. C. (2011). *On the dissociation between pitch production and pitch* *judgment: The role of musical expertise* (Unpublished master's thesis). University of Lethbridge.

Saenz, M., & Langers, D. R. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research*, *307*, 45-52.

Schellenberg, E., & Trehub, S. (2003). Good pitch memory is widespread. *Psychological Science*, *14*, 262-266.

Schneider, P., Dosch, H. G., Gutschalk, A., Scherg, M., Specht, H. J., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*, 688-694.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., . . . Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience*, *8*, 1241-1247.

Seither-Preisler, A., Krumbholz, K., Patterson, R., Johnson, L., & Nobbe, A. (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*, 743-751.

Selting, M. (1992). Prosody in conversational questions. *Journal of Pragmatics*, *17*, 315-345. Shamma, S. A. (2004). Topographic organization is essential for pitch perception. *PNAS*, *101*, 1114-1115.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, *36*, 2346-2353.

Siedenburg, K., & McAdams, S. (2017). Four distinctions for the auditory "wastebasket" of timbre. *Frontiers in Psychology*, *8*, 1747.

Singh, S. P. (2014). Magnetoencephalography: Basic principles. *Annals of Indian Academy of Neurology*, *17*, S107-S112.

Singing 101: Ranges and registers. (2017).

Smith, N., & Schmuckler, M. (2008). Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch possessors. *The Journal of the Acoustical Society of America*, *123*, 77-84.

Smith, S. W. (1999). The scientist and engineer's guide to digital signal processing. California Technical Publishing.

Song, X., Osmanski, M. S., Guo, Y., & Wang, X. (2016). Complex pitch perception mechanisms are shared by humans and a new world monkey. *Proceedings of the National Academy of Sciences*, *113*, 781-786.

Sonninen, A., Hurme, P., & Vilkman, E. (1992). Roentgenological observations on vocal fold length-changes with special reference to register transition and open/covered voice. *Scandinavian Journal of Logopedics and Phoniatrics*, *17*, 95-106.

Sound and music. (1997-2019). Retrieved from http://www.howmusicworks.org/103/Sound-and-Music/Amplitude-and-Frequency

Sound intensity. (n.d.). Retrieved from http://hyperphysics.phy-astr.gsu.edu/hbase/ Sound/intens.html

Stapel, D. A., & Semin, G. R. (2007). The magic spell of language: Linguistic categories and their perceptual consequences. *Journal of Personality and Social Psychology*, *93*, 23-33.

Stevens, S. S., & Newman, E. B. (1936). The localization of actual sources of sound. *The American Journal of Psychology*, *48*, 297-306.

Sussman, E., & Steinschneider, M. (2006). Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. *Brain Research*, *1075*, 165-174.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*, 345-361. Terhardt, E., Stoll, G., & Seewann, M. (1982). Pitch of complex signals according to virtual pitch theory: Tests, examples and predictions. *Journal of the Acoustical Society of America*, *71*, 671-678.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. (2004). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioural study. *Experimental Brain Research*, *161*, 1-10.

Thurlow, W. R. (1963). Perception of low auditory pitch: A multicue, mediation theory. *Psychological Review*, *70*, 461-470.

Thurlow, W. R., & Hartman, T. F. (1959). The "missing fundamental" and related pitch effects. *Perceptual Motor Skills*, *9*, 315-324.

Titze, I. (2004). *Principals of voice production*. National Center for Voice and Speech. *Tone language.* (n.d.). Retrieved August 2018, from https://dictionary.cambridge.org/dictionary/english/tone-language

Truax, B. (Ed.). (1978). *Handbook for acoustic ecology*. A.R.C. Publications, Vancouver. Truax, B. (Ed.). (1999). *Handbook for acoustic ecology*. Cambridge Street Publishing.

van Heuven, V. J., & van Zanten, E. (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication*, *47*, 87-99.

Vomberg, L. H. (2017, April). *An indirect measurement of pitch perception* [Unpublished Honour's Thesis].

Welch, G. F. (1979). Poor pitch singing: A review of the literature. *Psychology of Music*, *7*, 50-58.

Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, *9*, 519-527.

Woo, M. (2013, December). *Is valley girl speak, like, on the rise?* Retrieved August 2018, from https://news.nationalgeographic.com/news/is-valley-girl -speak--like--on-the-rise-/

Youngdahl, J. (2018, July). *Private communication.*

Zainal, M. R. M., Hussain, A., & Samad, S. A. (2012). *Recent progress in data engineering and internet technology* (Vol. 1; F. L. Gaol, B. N. University, P. Menteng, Jakarta, & Indonesia, Eds.). Springer.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *The Journal of the Acoustical Society of America*, *84*, 566-572.

Zatorre, R. J. (2003). Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience*, *6*, 692-695.