Factors affecting relative pitch perception

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Psychology

by

Carolyn Marie McClaskey

Dissertation Committee:
Professor Kourosh Saberi, Chair
Professor Gregory S Hickok
Associate Professor Amy M. Bauer

2016
DEDICATION

To mom, dad, and my brother Ian,
for everything.

To friends and family both near and far,
for your patience and support.

And to Ted, Travis, and Captain,
for your encouragement and love.
TABLE OF CONTENTS

LIST OF FIGURES iv
LIST OF TABLES v
ACKNOWLEDGMENTS vi
CURRICULUM VITAE vii
ABSTRACT OF THE DISSERTATION xi
INTRODUCTION 1

CHAPTER 1: Frequency and Pitch 3
A Brief Overview of the Auditory System 3
Theories of Pitch Perception 9
Relative Pitch and Objectives of Present Study 16

CHAPTER 2: Factors affecting the identification of frequency sweep direction 19
Mechanisms of FM and Frequency Sweep Perception 20
Goals and Approach of the Chapter 25
Experiment 1 26
Experiment 2 31
General Discussion 36

CHAPTER 3: Standard-interval size affect interval-discrimination thresholds 40
Methods 44
Results 49
Discussion 53
Conclusions 57

CHAPTER 4: Effects of frequency region on interval discrimination thresholds 58
Methods 62
Results 65
Discussion 68

CHAPTER 5: Summary and Conclusions 71
REFERENCES 74
# LIST OF FIGURES

| Figure 1.1 | An artist’s rendering of the ear | 4 |
| Figure 1.2 | Traveling wave patterns | 6 |
| Figure 1.3 | Phase-locking in the auditory pathway | 8 |
| Figure 1.4 | Representations of a harmonic complex tone with a F0 of 440 Hz | 11 |
| Figure 2.1 | Direction identification sensitivity at 500 Hz | 29 |
| Figure 2.2 | Predicted d’ performance at 500 Hz | 30 |
| Figure 2.3 | Direction identification performance at 8 kHz | 34 |
| Figure 2.4 | Predicted d’ performance at 8 kHz | 35 |
| Figure 3.1 | Sample stimuli in a pitch-interval task | 46 |
| Figure 3.2 | Grand averaged thresholds for each subject group | 49 |
| Figure 3.3 | Mean group thresholds as a function of standard-interval sizes | 51 |
| Figure 3.4 | Pitch-interval discrimination thresholds expressed as Weber Fractions | 53 |
| Figure 4.1 | Interval-discrimination performance by frequency region | 66 |
| Figure 4.2 | Boxplot of median thresholds by group in both frequency regions | 67 |
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Binned values for stimulus parameters of rate and transition span</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Interval-Discrimination Thresholds expressed as Weber Fractions</td>
<td>52</td>
</tr>
</tbody>
</table>
I would like to thank my committee chair, Professor Kourosh Saberi, whose humor and mentorship has been invaluable to me these past several years. I would like to thank my committee members, Professor Greg Hickok and Professor Amy Bauer, who have provided valuable advice and expertise at crucial moments throughout my graduate career.

I would also like to express my sincerest gratitude to my mentor and friend, Professor Barbara Sarnecka, whose passion for writing has given me the tools to succeed in academia and made this dissertation possible.

I thank Marcel Dekker Incorporated for permission to include Figure 1.1 as part of my dissertation, originally published in Hearing: An Introduction to Psychological and Physiological Acoustics (5th Ed.) by Stanley Gelfand (2010). I also thank AIP Publishing LLC and the Acoustical Society of America for permission to include Figure 1.2 of my dissertation, which was originally published in The Journal of the Acoustical Society of America. I furthermore thank both Springer Verlag for permission to include Figure 1.3 of my dissertation, originally published in the Springer Handbook of Auditory Research (SHAR) volume Pitch: Neural Coding and Perception, and the Society of Neuroscience for permission to include Figure 1.4 of my dissertation, which was originally published in The Journal of Neuroscience.

Financial support was provided by the University of California, Irvine, and an NIH T32 pre-doctoral fellowship from the UCI Center for Hearing Research.
CURRICULUM VITAE

Carolyn M McClaskey

2009  B.A. Music  
  *Boston College, Chestnut Hill, MA*

2009  B.S. Biochemistry, Honors Program  
  *Boston College, Chestnut Hill, MA*

2014  M.S. Cognitive Neuroscience  
  *University of California, Irvine, CA*

2016  Ph.D. Psychology, Cognitive Neuroscience Concentration  
  *University of California, Irvine, CA*

ACADEMIC EMPLOYMENT

2006–2008  Summer Research Assistant, Yale Child Health Research Center  
  PIs: Dr. Scott A Rivkees and Dr. Christopher C Wendler  
  *Yale School of Medicine, Yale University*

2009–2010  Research Assistant, Yale Child Health Research Center  
  PIs: Dr. Scott A Rivkees and Dr. Christopher C Wendler  
  *Yale School of Medicine, Yale University*

2010-2013  Graduate Teaching Assistant  
  *School of Social Sciences, University of California, Irvine*

2015  Course Instructor  
  Psych 9B: Introduction to Psychology  
  *Summer Session, University of California, Irvine*

2015-2016  Graduate Writing Consultant  
  *Graduate Resource Center, University of California, Irvine*

2010-2016  Graduate Student Researcher, Auditory Neuroscience & Language Lab  
  PIs: Dr. Kourosh Saberi (direct supervisor) and Dr. Greg Hickok  
  *Department of Cognitive Sciences, University of California, Irvine*


**PAPERS UNDER REVIEW / IN PREPARATION**


McClaskey, C. M. Musicians and nonmusicians may use different listening strategies in a pitch-interval discrimination task. (in preparation).

McClaskey, C. M., Saberi, K. A potential temporal mechanism in the perception of logarithmic frequency sweeps. (in preparation).

**AWARDS AND HONORS**

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan. 2011</td>
<td>Associate Dean’s Fellowship for Academic Excellence</td>
</tr>
<tr>
<td>June 2013</td>
<td>UCI Center for Hearing Research NIH T32 Pre-Doctoral Fellowship, $35,000 annually</td>
</tr>
<tr>
<td>Nov. 2013</td>
<td>Student Travel Award, Acoustical Society for America</td>
</tr>
<tr>
<td>June 2014</td>
<td>Dean’s Reading Group Small Grant: Cognitive Neuroscience of Meditation</td>
</tr>
<tr>
<td>June 2014</td>
<td>UCI Center for Hearing Research NIH T32 Pre-Doctoral Fellowship, $35,000 annually, Renewed (2-year maximum)</td>
</tr>
<tr>
<td>June 2015</td>
<td>Dean’s Reading Group Small Grant: The Cognitive Science Writing Group</td>
</tr>
<tr>
<td>June 2015</td>
<td>Falmagne Research Award, UCI Dept. of Cognitive Sciences, $3,000</td>
</tr>
</tbody>
</table>

**SERVICE**


2014 – 2015 Co-organizer and Host, CHR Seminar Speaker Series, UCI Center for Hearing Research, Irvine CA. Hosted speakers include Dr. Christophe Micheyl (Starkey), and Dr. Frederick J. Gallun (OHSU). Faculty advisor: John Middlebrooks.
TEACHING EXPERIENCE

A. Courses at the University of California, Irvine

Fall 2010  Teaching Assistant, Linguistics: Introduction to Phonology
Spring 2011  Teaching Assistant, Psychology 9B: Psychology Fundamentals
Fall 2011, 2012  Teaching Assistant, Psychology 112A: Experimental Psychology
Winter 2012, 2013  Teaching Assistant, Psychology 112B: Advanced Experimental Psychology
Spring 2012, 2013  Teaching Assistant, Psychology 112C: Research in Experimental Psychology
Summer I 2014  Student Grader, Psychology 9B: Psychology Fundamentals
Summer II 2014  Teaching Assistant, Brain Disorders
Summer II 2014  Teaching Assistant, Psychology 9B: Psychology Fundamentals
Summer I 2015  Course Instructor, Psychology 9B: Psychology Fundamentals
Summer II 2015  Course Instructor, Psychology 9B: Psychology Fundamentals

PROFESSIONAL MEMBERSHIPS

Society for the Neurobiology of Language
Association for Research in Otolaryngology
Acoustical Society of America

TECHNICAL SKILLS

Notable Soft Skills: Excellent verbal/written and communication, exceptional writing skills, effective resource/time management, resourceful problem-solving, strong interpersonal skills, mentoring

Computer Skills: MATLAB, SPSS, Excel, Adobe Photoshop

Scientific Skills: psychophysics, electroencephalography (EEG), signal processing, acoustic measurements, statistics, OAEs, spectroscopy, immunohistochemistry, cell culture, tissue-dissection, cryo-sectioning, fluorescent microscopy, PCR & real-time PCR, protein isolation, Western Blotting

CERTIFICATIONS

Course Design Certificate Workshop
Center for Engaged Instruction, February 2016, University of California, Irvine

Sound Level Meter Users Group
Brüel & Kjaer, Sound and Vibration Conference, April 2016, Long Beach, CA
ABSTRACT OF THE DISSERTATION

Factors affecting relative pitch perception

By

Carolyn M McClaskey

Doctor of Philosophy in Psychology

University of California, Irvine, 2016

Professor Kourosh Saberi, Chair

Sounds that evoke a sense of pitch are ubiquitous in our environment and important for speech, music, and auditory scene analysis. The frequencies of these sounds rarely remain constant, however, and the direction and extent of pitch change is often more important than the exact pitches themselves. This dissertation examines the mechanisms underlying how we perceive relative pitch distance, focusing on two types of stimuli: continuous pitch changes and discrete pitch changes.

In a series of experiments testing continuous pitch changes, listeners heard pure-tone frequency sweeps and reported whether they moved up or down. Sweeps varied in the extent of frequency change, the rate of frequency change, and sweep center frequency. Results provide evidence for a sampling mechanism in which listeners extract the start and end pitches of each sweep and then compare them to determine sweep direction. A comparison of performance between frequency regions shows a smaller effect of sweep rate at high frequencies (>6 kHz), suggesting that the mechanism by which listeners extract start/end pitches at low frequencies is based on temporal pitch processing.
To examine discrete frequency changes, nonmusicians, amateur musicians, and formally trained “expert” musicians heard two different pitch-intervals and were asked to indicate which was larger. Intervals varied in the size of the comparison interval and were presented in both low and high frequency regions. Expert musicians performed significantly better than other listeners, while amateur musicians performed similar to nonmusicians. Contrary to previous studies, all groups demonstrated better performance for smaller intervals. A comparison of frequency region also suggests a potential difference in listening strategy between groups: nonmusicians produced higher thresholds at high frequencies but amateur and expert musicians did not.

Overall, results provide novel evidence for the role of a sampling mechanism in sweep-direction identification, and present a previously undocumented effect of standard interval size in pitch-interval perception. The effects of frequency region found in both contexts furthermore suggests that temporal pitch processing mechanisms are used at low frequencies, and that different listening strategies may be used for relative pitch perception at higher frequencies where temporal pitch cues are less reliable.
INTRODUCTION

Our environment contains a cacophonous mixture of overlapping sound waves generated by mechanically vibrating sources all around us, but these sounds reach us as only a single sound pressure wave in each ear. The monumental task of our auditory system is to separate out these overlapping sounds signals and build higher-order representations that help us interact with and navigate through our environment. This is a difficult task that requires information about the waveforms, component frequencies, source location, timing, and energy content of the auditory signals, and is sometimes only possible with prior knowledge of the environment or with information from our other senses.

When we interpret the physical properties of a sound stimulus, one of the most important features is frequency content. This dissertation focuses on the dimension of frequency, and how we perceive relationships between sounds of different frequencies. This type of perception, more broadly known as relative pitch perception, underlies many aspects of hearing, including speech perception, music perception, and spatial hearing.

The broader topic of frequency perception inherently involves a discussion of the field’s two complementary theories of spectral and temporal pitch mechanisms, which are presented in Chapter 1. Chapter 1 first begins with a brief introduction to the relevant parts of the auditory system – most importantly the cochlea and the auditory nerve – and then provides a brief overview of these theories of pitch perception and the basic properties of each. Following this, I present a series of original experiments of relative pitch perception that focus specifically on factors affecting relative pitch perception and how temporal processing contributes to our perception of relationships between frequencies. Chapter 2
presents two experiments that investigate how we perceive the direction of pitch change in dynamically changing frequency, *i.e.* frequency sweeps, and discusses how the findings provide support for sampling mechanisms based on temporal pitch processing in this context. Chapters 3 and 4 investigate the perception of discrete changes in frequency, called *pitch-intervals*. Chapter 3 highlights an important but previously undocumented effect of standard interval size in pitch-interval discrimination, while Chapter 4 examines how this type of perception may also rely on temporal pitch processing and how it is affected by musical experience. Chapter 5 concludes the dissertation with a summary of the main findings.
CHAPTER 1

Stimulus frequency is one of the primary physical properties of sound. Frequency is defined as the rate of repetition of a periodic sound wave, and is quantified as the number of cycles per second, or Hertz. When we process the frequency of a sound wave, what we perceive is pitch. In other words, pitch is the perceptual correlate of the frequency of periodic sounds, and is officially defined as “that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high” (ANSI, 1994).

When we hear a sound double in frequency, we hear a linear increase in pitch. Pitch is not therefore not associated with units of Hertz, but is instead quantified in units of octaves and semitones. An octave is defined as the distance between a tone and one double its frequency, and octave distances are defined on a logarithmic scale as:

\[
x = \log_2 \left( \frac{f_1}{f_2} \right)
\]

or as:

\[
f_1 = f_2 2^x
\]

where \(f_1\) and \(f_2\) are the two frequencies of interest and \(x\) is their distance in octaves. A semitone is defined as one-twelfth of an octave, and is equivalent to a “half-step” in the most common Western musical theory tuning system, 12-tone equal-tempered tuning (12-TET). Semitones and octaves thus provide a useful way of equating perceptual pitch distance across frequencies.

A brief overview of the auditory system

The extraction of a pitch percept from a periodic stimulus involves all parts of the auditory system, beginning at auditory periphery and extending through the auditory
nerve and the brainstem to the cortex. While a full discussion of the auditory system is outside the scope of this dissertation, the basic phenomena and properties of the cochlea and the auditory nerve are relevant to future chapters and are briefly presented here.

Figure 1.1. An artist’s rendering of the ear. From Gelfand (2010) with permission.

The auditory periphery has three parts (Figure 1.1): the outer ear, middle ear, and inner ear. The outer ear includes the pinna and the concha (the visible ear cartilage), the external ear canal, and the ear drum, also called the tympanic membrane (Fuchs, 2010; Pickles, 2013). The middle ear contains the human body’s three smallest bones, the malleus, incus, and stapes, which are located one after the other in that order just inside the tympanic membrane. When a sound pressure wave reaches our ear, it gets filtered by the shape of the pinna and travels down our ear canal to vibrate the tympanic membrane. The middle ear bones then pick up these vibrations and transfer them to the inner ear, which
contains the *cochlea*; in doing so the middle ear also converts the sound waves from air pressure waves to fluid pressure waves (Fuchs, 2010).

The cochlea is a coiled tube with three parallel fluid-filled compartments separated by membranes, which span the length of the cochlear duct (Fuchs, 2010; Robles & Ruggero, 2001; Slepecky, 1996). Sound vibrations enter the base of the cochlea when the stapes makes contact with the oval window. These waves then travel the length of the cochlea, pass through the helicotrema at the apex, and travel back down the duct (through a separate membranous compartment) to disperse through the round window. Fluid pressure waves passing up the cochlea deflect the cochlear partition, which consists of the basilar membrane, the Organ of Corti, and the outer and inner hair cells that are anchored in the Organ of Corti (Slepecky, 1996). The basilar membrane is a thin gelatinous membrane located just underneath the Organ of Corti, and which is stiff and narrow at its base, and wider and flexible at the cochlear apex (von Békésy, 1960). When a sound wave propagates through the cochlear fluid, it displaces the basilar membrane in the form of a traveling wave that peaks at different places on the basilar membrane according to the sound’s frequency components (Figure 1.2).
Figure 1.2. Traveling wave patterns. (a) executed by a hypothetical ribbon-like partition; (b) observed along a single-layer partition, a hypothetical basilar membrane. The left side of the diagram indicates the basal side of the basilar membrane and the right side corresponds to the apical side of the basilar membrane. The traveling wave moves left to right. Scales are arbitrary in both drawings. Reprinted with permission from Tonndorf (1960). Copyright (1960) The Acoustical Society of America.

The narrow base of the basilar membrane is preferentially activated by frequency with short wavelengths (i.e. high frequencies) while the wider apical end responds maximally to waves of longer wavelengths and lower frequencies (Greenwood, 1961; von Békésy, 1960). Inner and outer hair cells arrayed along the length of the membrane are perturbed by this displacement, and fire when moved. Thus, the properties of the basilar membrane and the cochlea introduce a mechanism by which mechanical waves become transduced into neuronal signals.

The frequency selectivity of the basilar membrane underlies many aspects of hearing, and the frequency gradient it introduces – defined as tonotopic organization – is preserved at all subsequent levels of the auditory system (Guinan, Norris, & Guinan, 1972;
Humphries, Liebenthal, & Binder, 2010; McCreery, Shannon, Moore, & Chatterjee, 1998; Merzenich & Reid, 1974; Pantev et al., 1988, 1995; Rubel & Parks, 1975). Because of this organization and its properties, the cochlea is often conceptualized as a bank of overlapping band-pass frequency filters, each with slightly different center frequencies as well as asymmetric shapes that are a result of the shape of the basilar membrane’s mechanical response (as shown in Figure 1.2) (Glasberg & Moore, 1990; Patterson, Allerhand, & Giguère, 1995). Referring to the cochlea in this way, often with the terms auditory filters or the auditory filterbank, is a mathematically useful way of representing its functional properties and is ubiquitous throughout both quantitative and qualitative models of the auditory system (de Cheveigné, 2005; Moore, 2012).

Once hair cells are activated by the displacement of the basilar membrane, their electrical signal is transmitted to the VIII cranial nerve, called the auditory nerve (AN). The AN is comprised of thousands of nerve fibers that are commonly divided into 3 groups based on their spontaneous discharge rate (SR): low-, medium-, and high-SR fibers (Liberman, 1978). One of the most important properties of auditory nerve fibers is their propensity to synchronize their electrical discharges to amplitude peaks in the stimulus waveform (Galambos & Davis, 1943). In other words, they fire at a specific phase of the waveform, a process known as phase-locking (See Figure 1.3).
Figure 1.3. Phase-locking in the auditory pathway. (A) Action potentials of a neuron that is firing in response to the amplitude peaks of a 300 Hz pure tone. (B) Stimulus waveform, 300 Hz. (C) Vector strength of phase-locking as a function of frequency for the cat (Johnson, 1980), guinea pig (Palmer & Russell, 1986), and barn owl (Köppl, 1997). Note how the upper limit of phase-locking varies across species. From Winter (2005) with permission.

Auditory-nerve phase-locking is not possible at all stimulus frequencies, however, and begins to deteriorate at higher frequencies, likely as a result of the low-pass filtering of the inner-hair cells and synapses (Palmer & Russell, 1986; Weiss & Rose, 1988), and specific characteristics of cells (see Joris & Verschooten, 2013). The point at which phase-locking and neural synchrony begins to decline is known as the phase-locking cut-off frequency, and the maximum frequency at which phase-locking is possible is commonly called the upper limit of phase-locking. The precise values of the cut-off frequency and upper limit vary across species (Johnson, 1980; Köppl, 1997; Palmer & Russell, 1986; Rose,
Brugge, Anderson, & Hind, 1967) (see Figure 1.3c) and are unknown in humans. However, because many psychophysical and physiological data in humans indicate a shift in our perception between 4 and 5 kHz, it is generally accepted that the upper limit of AN phase-locking in humans is 4-5 kHz, although some have suggested it may be as high as 6 kHz (Oxenham, Micheyl, Keebler, Loper, & Santurette, 2011). The proposed cut-off frequency in humans is also unknown, but may be around 1-2 kHz (Joris & Verschooten, 2013; Verschooten, Robles, & Joris, 2015; Winter, 2005). While there is considerable debate about the exact limits of phase-locking in humans (Joris & Verschooten, 2013; Moore & Sek, 2009; Oxenham et al., 2011; Sek & Bukała, 2013), it’s nevertheless clear that phase-locking is robust and reliable below a certain frequency, above which it begins to degrade as stimulus frequency increases until it is eventually completely unavailable. Phase-locked neurons are present at all levels of the auditory system and are important for many aspects of hearing such as speech-in-noise processing and spatial hearing, but are particularly relevant to theories of pitch perception.

**Theories of pitch perception**

The process by which we extract pitch from periodic sound stimuli involves the integration of two complementary mechanisms: a spectral mechanism and a temporal one. Spectral mechanisms focus on the spectral distribution of energy in a signal, and temporal mechanisms emphasize the temporal and periodic qualities of the stimulus waveform; As Alain de Cheveigné says, “One involves the spectrum and the other the waveform” (de Cheveigné, 2005).
Rate-Place theories of pitch perception

Spectral pitch theories posit that the auditory system extracts a pitch based on how energy is distributed across frequencies, which is communicated by the place of maximum displacement of the basilar membrane. Information about the degree of such displacement is conveyed via the firing rate of auditory nerve fibers: the greater the membrane displacement, the faster the neurons at that place fire. At high sound levels (60-70 dB), this firing rate information is most likely conveyed by low-SR fibers, while high-SR fibers likely communicate this information at low sound levels (less than 50 dB) (Fuchs, 2010; Pickles, 2013).

In the case of pure tones, or simple sinusoidal waveforms that contain energy at only a single frequency (as in Figure 1.3b), the traveling wave peaks at only one frequency, and its location indicates the pitch that we hear (Plack & Oxenham, 2005; Plack, Oxenham, Fay, & Popper, 2005). For more complex tones that contain many frequencies, often the pattern of excitation determines our pitch percept (Oxenham, 2012; Plack & Oxenham, 2005; see Figure 1.4). For this reason, this mechanism is also commonly referred to as place coding, or rate-place coding, and the cues that give rise to this information are called energy cues or excitation-pattern cues.

There are several important features of spectral pitch theories, which arise from the specific properties of the basilar membrane. First, rate-place coding is very fast, and requires very short temporal integration times (Moore, 2012; Pickles, 2013). Second, these types of energy cues are available across the entire length of the basilar membrane and for frequencies across the entire range of human hearing, which is between 20 and 20,000 Hz.
Figure 1.4. Representations of a harmonic complex tone with a F0 of 440 Hz. A, Time waveform. B, Power spectrum of the same waveform. C, Auditory filter bank, representing the filtering that occurs in the cochlea. D, Excitation pattern, or the time-averaged output of the auditory filters. E, Sample time waveforms at the output of the filter bank, including filters centered at the F0 (440 Hz), 4th harmonic (1760 Hz), 8th harmonic (3520 Hz), and 12th harmonic (5280 Hz) of the complex. The 4th harmonic is resolved because only one harmonic exists within a given auditory filter. At higher harmonics numbers, auditory filters are wider yielding unresolved harmonics. Figure from Oxenham (2012) with permission.
Because basilar membrane displacement is inherently a measure of energy at any
given frequency, rate-place coding is also easily disrupted by extraneous variations in
energy that might not have anything to do with the stimulus frequency (Dooley & Moore,
1988; Moore & Sek, 1998). Auditory filters also broaden with increasing level, especially on
the low-frequency side (Pickles, 2013), and shifts in sound level have been shown to move
the peak of a traveling wave by as much as half an octave at certain frequencies (McFadden,
1986). This energy dependence is evident in psychophysical data, which show that the
pitch of pure tones is dependent on sound level (Stevens, 1935; Verschuure & Van
Meeteren, 1975): increases in sound level lead to decreases in pitch for tones below 2000
Hz and for tones shorter than 40ms (Houtsma & Rossing, 1987), while increases in sound
level lead to increases in pitch for tones above 2000 Hz. For a more in-depth discussion of
psychophysical evidence for rate-place coding, see Plack et al., (2005) and Moore (2012).

Lastly, because the membrane moves as a wave and auditory filters are not infinitely
narrow – i.e. our frequency resolution is not perfect – any change in the frequency of a tone
must result in a noticeable change in the peak of excitation for it to be heard. Zwicker
(1970) developed a model of frequency discrimination, which proposed that a 1dB change
in excitation at the output of an auditory filter would be enough to detect a change in pitch.
However, psychophysical data show much better performance than would be predicted by
these and other place models of frequency detection and frequency-modulation detection
(Moore, 1973, 1974; Moore & Sek, 1996), particularly at stimulus frequencies below 4 kHz.
Models of pitch perception have attempted to reconcile this difference and explain our
exceptional pitch perception ability by suggesting that we use information from both sides
of an auditory filter (Moore & Sek, 1998) or all filters together (Moore & Sek, 1992, 1994),
but this has proven elusive (for a more in-depth discussion, see B. C. J. Moore, 2012; Plack & Oxenham, 2005).

There is extensive evidence for the role of rate-place coding in pitch perception, and it is thought to be the primary mechanism for extracting of the pitch of pure tones and the pitch of complex tones with unresolved harmonics\(^1\) (Gockel, Moore, Carlyon, & Plack, 2007; Marmel et al., 2015; Moore, 1973, 2012; Plack & Oxenham, 2005). However, there exist other pitch phenomena that spectral theories are unable to explain. For example, stimuli can elicit a pitch even if there is no energy at the frequency corresponding to the pitch percept, which happens in the case of the missing fundamental, also called residue pitch (Schouten, 1940). White noise can also elicit a pitch if it has been manipulated so that its waveform has sufficient periodicity, as is the case for sinusoidally amplitude modulated (SAM) noise and iterated rippled noise (Burns & Viemeister, 1976, 1981; Yost, 1996). In such cases, the pitch that we hear does not correspond to the peak of energy in the signal – since there is no peak for white noise – and spectral theories of pitch fall short in explaining this type of pitch. Expanded rate-place theories have attempted to explain such phenomena using features such as template matching (Goldstein, 1973; Terhardt, 1974; Wightman, 1973), but have not been entirely successful. Instead, researchers turn to a second class of pitch theories.

---

\(^1\) A harmonic complex sound has multiple frequency components at integer multiples of a common fundamental frequency, \(F_0\). For example, a complex harmonic tone with a fundamental frequency of 440 Hz would have its first harmonic at a frequency of 880 Hz, its second harmonic at a frequency of 1320 Hz, and so on. Resolved harmonics are lower integer multiples (usually <7), and are called resolved because the auditory filter activated by that harmonic does not contain any other harmonics. A visual representation can be seen in Figure
Temporal theories of pitch perception

This alternate class of theories is collectively known as temporal theories. These generally rely on the periodicity of the waveform itself, which is represented in the temporal pattern of phase-locked auditory nerve fibers. Because phase-locked auditory nerve fibers fire at a specific phase of the stimulus waveform, the AN population response contains a representation of the periodicity of the waveform. This population response is thought to be decoded at higher levels, generally via an autocorrelation analysis (Cariani, 1999; Cariani & Delgutte, 1996a; Meddis et al., 2013; Voutsas, Langner, Adamy, & Ochse, 2005), and thus the temporal periodicity of the waveform gives rise to the pitch that we hear.

Because the non-random discharge patterns of the AN fibers are able to accurately represent the exact timing of the peaks in the waveform at sufficiently low waveform repetition rates, temporal pitch can be very precise (Heinz, Colburn, & Carney, 2001; Heinz et al., 2001; Moore, 1973; Siebert, 1970) and models using temporal coding are able to accurately explain our low-frequency discrimination thresholds, and the increase of frequency different limens (FDLs) with increasing stimulus frequency that spectral theories cannot explain. Temporal theories can also explain many other pitch phenomena, such as the missing fundamental and periodicity pitch (Cariani & Delgutte, 1996b), because they focus on the waveform periodicity rather than spectral energy.

This type of mechanism also has limits, however, which arise primarily from the properties of phase-locking discussed previously, namely that phase-locking can only represent periodicities of sufficiently slow rates. This also results in a second consideration: a temporal mechanism requires a longer integration window in order to
generate a stable percept. In other words, if the stimulus is too short or has too few
repetitions, then the system doesn’t have enough time to determine the repetition rate, i.e.
frequency, of the stimulus. As a result, temporal pitch processing is more adversely affected
by short-duration stimuli than are rate-place mechanisms. In sum, temporal pitch is
generally limited to frequencies below about 4kHz, and is best for stimuli of at least 80 ms,
or even longer (~100ms) if stimuli are low frequency (White & Plack, 1998). For this
reason, the temporal system is often called “sluggish” (Moore, 2012; Moore & Sek, 1995,
1996; Sek & Moore, 1995).

For several reasons, temporal pitch processing is thought to underlie musical pitch.
First, melodies cannot be constructed with stimuli above 4-5 kHz (Attneave & Olson, 1971;
Ward, 1954; but see Oxenham et al., 2011), e.g. the highest note of a musical instrument is
approximately 4.5 kHz. Second, musical pitch perception is degraded in situations where
phase-locking to the temporal fine structure of a signal is unavailable, as is the case with
cochlear implants (Penninger et al., 2013). Lastly, melodies and harmonies can be reliably
constructed using rippled noise stimuli (Yost, 1996), which is a specific type of stimuli in
which spectral cues are unavailable but stimulus periodicity is maintained.

Like spectral theories, temporal theories alone are unable to explain all pitch
phenomena; pitch perception in the real world is likely a combination of the two, and the
use of one versus the other may depend on context, stimuli, and the individual strengths of
the listener. A large body of psychophysical evidence supports this idea, and shows a shift
in perception between approximately 4000 Hz and 5000 Hz. For example, frequency
discrimination performance (quantified via the FDL) becomes markedly worse at
frequencies above 4000 Hz (Moore, 1973; Sek & Moore, 1995). The effect of duration on
FDLs also changes at approximately 4000 Hz: below this point, increasing stimulus duration leads to better discrimination performance and lower FDLs, while this effect decreases as frequency increases (Plack & Oxenham, 2005). Random variations in sound level – which disrupts energy cues but not phase-locking cues – also have a very large effect on frequency discrimination for tones above 4000 Hz, but less of an effect for lower frequency tones. Place models of frequency discrimination and frequency modulation detection also fall apart for stimuli below 5000 Hz. Taken together, these results have led to the broad conclusion that temporal pitch likely dominates at frequency below approximately 5 kHz, and rate-place coding dominates at higher frequencies (Moore & Sek, 1996; Oxenham, 2012; Pickles, 2013; Plack & Oxenham, 2005; Sek & Moore, 1995).

**Relative pitch and objectives of the current study**

Pitch sounds are ubiquitous in our environment. In processing these stimuli, however, often the relationships between pitches is just as informative as the pitches themselves, and our ability to perceive these relationships is known as a sense of *relative pitch*. Relative pitch exists at several levels and can be specific enough to quantify the exact semitone distance separating two sounds, or only general enough to identify one as being higher than the another – in its most basic form it determines whether two pitches are the same or different. Such percepts help us differentiate between the vocal intonation patterns of a sincere question and a sarcastic remark, and between the tonal qualities of a Major key and a Minor one in music. Relative pitch perception is therefore an important function of the auditory system and a fundamental aspect of both speech and music perception.
In this dissertation, I examine the mechanisms underlying relative pitch perception and the factors that influence this perception, focusing on two types of stimuli that are important in many aspects of hearing, particularly speech and music: continuous pitch changes (frequency sweeps) and discrete pitch changes (intervals). Because pitch—musical pitch in particular—is thought to be primarily temporal based but the role of temporal theories is less often studied in relative pitch perception, I focus in particular on the role of temporal processing and the effect of phase-locking cues. Furthermore, because this may be affected by auditory experience, I also examine how this ability differs between individuals with auditory experience, namely, musical training.

In sum the studies presented here address the broader experimental context of how we make judgments about relationships between frequencies, focusing specifically on the primary experimental questions of:

1. What are the factors affecting our perception of (a) pitch change direction in logarithmic frequency sweeps, and (b) pitch distance in melodic pitch-intervals?
2. Are temporal phase-locking cues involved in the relative pitch perception?
3. How does relative pitch perception differ between individuals with different types of auditory experience?

In Chapter 2 I focus specifically on the perception of direction in frequency sweeps, a paradigm that is less often studied and which may recruit a specific type of temporal-pitch processing mechanism, a *sampling mechanism*. This mechanism was originally proposed by Hartmann and Klein (1980) and has been applied to many other related

---

2 In this dissertation, I focus primarily on the pitch of pure tone stimuli, but a broader discussion of pitch generated by other kinds of stimuli found in Oxenham and Plack (2010) and Moore (2012).
contexts of relative pitch perception, but has not been directly investigated in direction identification (a percept critical for tonal language perception). In Chapters 3 and 4 I investigate factors affecting the perception of pitch distance for discrete changes in pitch, called *pitch-intervals*. Chapter 3 examines the effects of standard-interval magnitude on nonmusicians, amateur musicians, and expert musicians, while Chapter 4 extends these results to include the effects of frequency region, a manipulation which takes advantage of the frequency limitations of phase-locking and which is often used as a psychophysical way of evaluating temporal processing. Both chapters also investigate how pitch-interval perception differs between individuals with and without musical training. Chapter 5 ends the dissertation by summarizing these findings and situating them within the broader context of relative pitch perception.
CHAPTER 2:

Factors affecting the identification of frequency sweep direction

Although pitch in general is an important feature of sound and is fundamental to music, language, and auditory scene analysis, it is dynamically changing pitch that is especially important for spoken communication. We most commonly encounter dynamically changing pitch in the form of frequency modulated (FM) sounds, in which frequency of the signal modulates back and forth, or as frequency sweeps, which are sounds that sweep either up or down in frequency. Frequency sweeps are particularly important for communication. In tonal languages, we use frequency sweeps of different directions to indicate semantic content: a word spoken in a rising pitch contour has different meaning than the same word spoken with a falling pitch contour (Y. Ye & Connine, 1999). In non-tonal languages, pitch contours and vocal pitch levels add emotional valence, arousal, and other nonverbal meaning to speech in non-tonal languages (Bänziger & Scherer, 2005; Grichkovtsova, Morel, & Lacheret, 2012; Scherer, 2003; Scherer, Ladd, & Silverman, 1984). While the interpretation of vocal emotion, lexical tones, and non-verbal intent is affected by a multitude of factors including context, semantic content, and cultural influences, it relies fundamentally on a listener’s ability to detect features of these basic auditory cues - namely, the direction of pitch change, the rate at which the pitch is changing, and the extent of change. The current chapter focuses primarily on frequency sweeps and how listeners identify their direction (up versus down).
Mechanisms of FM and frequency sweep perception

Dynamic Mechanism

FM and frequency sweeps have been extensively studied in the auditory field, and several mechanisms have been implicated in their perception. Physiological studies in animals have identified cortical FM-processing neurons that show selective tuning to the direction and rate of FM sweeps (Gordon & O’Neill, 1998; Tian & Rauschecker, 1994, 2004; Zhang, Tan, & Schreiner, 2003). The velocity and directional selectivity of these cells is thought to arise from a combination of excitatory and inhibitory sideband inputs that respond to neighboring frequencies: the neuron fires if it receives input from its excitatory sideband before it receives input from its inhibitory sideband, a situation only caused by a sweep moving in the preferred direction at the preferred velocity (Brimijoin & O’Neill, 2005; Gittelman & Li, 2011; Gordon & O’Neill, 1998; Kuo & Wu, 2012; Zhang et al., 2003). Studies using magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) have provided evidence for this direction-selective mechanism in humans, but primarily focus on more rapid frequency sweeps greater than an octave per second, and sometimes as rapid as 25 octaves/per second (Gordon & Poeppel, 2002; Hsieh, Fillmore, Rong, Hickok, & Saberi, 2012; König, Sieluzycki, Simserides, Heil, & Scheich, 2008; Pardo & Sams, 1993; Sams & Näätänen, 1991).

Sampling mechanism

Human psychophysical studies of frequency sweep perception generally find support for a mechanism that encodes sweeps as a primary stimulus feature, which is typically referred to as a dynamic mechanism (Demany, Carlyon, & Semal, 2009; Dooley &
Moore, 1988; Lyzenga, Carlyon, & Moore, 2004), exemplified most clearly in the fact that a frequency difference between two tones is more easily identified if the tones are connected by a frequency sweep (Demany et al., 2009; Lyzenga et al., 2004; Sek & Moore, 1999).

However, these studies also highlight findings that cannot be explained by such a mechanism, especially for sweeps with rates of frequency change slower than 1 octave per second. In an investigation of a listener’s ability to detect slow FM sweeps that were bounded by plateaus – or periods of time where frequency remains steady – Tsumura, Sone, and Nimura (1973) found that plateau duration affected detection thresholds; the shorter the plateaus, the more difficult it was to detect a frequency transition (Tsumura, Sone, & Nimura, 1973). Dooley and Moore (1988) also found an effect of duration, with shorter durations leading to poorer performance, although this effect disappeared at durations greater than 100ms (Dooley & Moore, 1988). Similarly, Demany, Carlyon, and Semal (2009) found that a frequency difference between two tones is more difficult to identify if it is based on a continuous change in frequency versus a discrete change (Demany et al., 2009). Such effects of duration and velocity would not be expected from a dynamic mechanism that is sensitive to the sweep per se; in fact, a dynamic mechanism might predict the opposite pattern of results since directions-selective cells often response best to extremely fast sweeps and are not as affected by duration (Zhang et al., 2003). It is worth noting, however, that nearly all psychophysical studies mentioned here involve sweeps with relatively slow rates of change (e.g. less than an octave per second or even slower), while physiological studies of direction-selective cells often use sweeps that move at multiple octaves per second.
Taken together, these findings suggest the involvement of additional mechanisms underlying the perception of slowly changing frequency. These additional mechanisms take many forms, but the most common is one based on Hartmann and Klein’s (1980) model of FM detection for low-modulation rates (*i.e.* slowly changing frequency) (Demany & Semal, 1989; Hartmann & Klein, 1980). In this *sampling-differencing model*, also called the *sampling model* or a *snapshot model*, a listener “takes successive pitch samples and compares them in an attempt to detect a change” (Hartmann & Klein, 1980). Although Hartmann and Klein’s sampling model was originally proposed to explain FM detection, it has also been used to explain to sweep detection, discrimination, and most importantly, direction identification (Demany & Semal, 1989; Dooley & Moore, 1988; Moore & Sek, 1995; Sek & Moore, 1999). In the context of sweep direction identification, a listener determines the direction of frequency change by sampling the frequencies of the stimulus at various points and then comparing them to evaluate whether a sweep moved up or down. Although it’s feasible for listeners to compare points throughout the sweep if stimulus duration is long enough, the simplest case of this model assumes that listeners compare only the start and end frequencies, making this model conceptually similar to the one proposed to explain the discrimination of steady pulsed tones (Sek & Moore, 1995) and perception of auditory motion (Grantham, 1986).

The sampling mechanism has several key features that are worth noting. Because stimulus frequency cannot be instantaneously determined, each successive sample of the stimulus requires an integration window of a certain duration, thought to be about 25 ms (Demany & Semal, 1989). If stimulus duration is too short, the system must either shorten the duration of each sample or sample fewer time points – both of which adversely affect
the estimate of instantaneous frequency. Put differently, the model predicts that performance increases monotonically with stimulus duration. The system is similarly adversely affected by rate of frequency change: if stimuli frequency is rapidly changing within a sampling window, the point estimate of instantaneous frequency will be less precise because the instantaneous frequency changes too rapidly over the course of the sampling window, resulting in a noisy estimate and impairing performance. Thus, this sampling mechanism is affected by stimulus duration and rate of frequency change in a way that the dynamic mechanism is not.

In sampling the stimulus frequency at successive time points along the sweep's trajectory, a subsequent question arises: which peripheral cues are used to establish pitch estimates in each successive sample? On one hand, listeners may form a pitch percept using rate-place mechanisms and energy cues. These spectral cues are generated by monitoring the output of filters on the basilar membrane and using the place of maximum excitation to infer the frequencies of the sweep at points along its path, as discussed in Chapter 1. This approach is closely related to Zwicker's (1970) model of FM detection\(^3\), but is adapted to apply to direction identification.

Such spectral energy cues have several distinct advantages, as discussed in Chapter 1. There are two main advantages that are relevant to the current discussion: First, this type of mechanism acts very quickly, and requires a minimum stimulus duration of only about 25 ms (Gordon & Poeppel, 2002); Second, because auditory filters span the entire length of the basilar membrane, this type of mechanism operates over the entire range of

\(^3\) in which FM is detected if “it resulted in a change in the amount of excitation greater than 1 dB at any point on the excitation pattern” (Moore & Sek, 1998; Zwicker, 1970).
human hearing and is not limited to low frequencies as would a temporal mechanism. Studies of FM detection and discrimination have found evidence for a mechanism that recruits energy cues, showing that superimposed energy changes have a slight effect on performance (Dooley & Moore, 1988; Moore & Sek, 1998; Sek & Moore, 1999) and that performance is roughly equivalent across frequency regions (Madden & Fire, 1996, 1997). It is worth noting, however, that this evidence exists primarily in studies of sweep detection and discrimination, and less often for sweep direction perception. Furthermore, the results of many of these studies cannot be fully explained by purely spectral cues: there is in fact a slight effect of center frequency, and energy roves are less detrimental than predicted, leading many researchers to suggest that a second type of cue is used.

Once again, as a complement to energy cues, a second strategy for extracting the pitch of successive samples may instead rely on temporal processing, in which phase-locking is used to establish a temporal pitch percept at each successive time point. Although potentially able to generate a more precise estimate of the sweep frequency at a given time point, this strategy also has its limitations (see Chapter 1); because a certain number of stimulus cycles are required for such an estimate, the system requires a longer temporal integration time in order to generate a stable pitch estimate. In other words, pitch estimates from this system are adversely affected by inadequate integration windows, which might arise from short duration stimuli or stimuli with rapid rates of frequency change – similar to the sampling mechanism. Due to these similarities, it is generally accepted – though less often tested – that a sampling mechanism is based on phase-locking cues. This assumption has not yet been directly examined, however, allowing for one important possibility: a sampling mechanism may use temporal pitch processing to extract
the pitch of start/end samples at low frequencies, and may use a rate-place pitch coding to extract the pitch of samples at high frequencies. If this were the case, a sampling mechanism based on spectral cues would be less adversely affected by rapid rates than a sampling mechanism based on phase-locking cues.

Studies examining the use of these two complementary cues in FM detection generally find support for the use of phase-locking cues in the detection of frequency modulation and frequency sweeps for low-frequency stimuli (generally below about 2 kHz) and have slow rates of change or modulation rates below approximately 2 Hz (Moore & Sek, 1995, 1996). At faster rates of change, shorter durations, and higher frequencies, spectral cues may dominate (Madden & Fire, 1996, 1997, Moore & Sek, 1995, 1998; Sek & Moore, 1995). However, many of these studies require listeners to simply detect a change or discriminate between two sweeps of the same direction, and fewer studies have investigated the mechanisms underlying direction identification in particular. One recent study of direction identification finds that more rapid sweep rates are more difficult to identify (Gordon & Poeppel, 2002), but it did not specifically address the role of phase-locking.

**Goals and approach of the chapter**

Thus, the goal of the current study is to test the hypothesis that a sampling mechanism is used to identify the direction of slow frequency sweeps, and to evaluate whether this sampling model is based on phase-locking cues at low frequencies. Relatively few psychophysical studies have specifically investigated how listeners determine the direction of slow-moving frequency sweeps (Pollack, 1968; Sergeant & Harris, 1962), and
to our knowledge, none have thus far directly investigated the role of temporal cues and the effects of frequency region in this task.

To do this, we used a glide direction identification task in which subjects were presented with a single glide that moved either up or down in frequency, and asked subjects to determine its direction. We varied the parameters of rate and transition span, and hypothesized that faster rates (and thus shorter durations) will negatively affect performance because this degrades the resolution of a sampling mechanism’s snapshots. Furthermore, we predicted that small transition spans would also negatively affect performance. To directly evaluate the role of temporal cues in this task, Experiment 1 tested subjects using sweeps with center frequencies near 500 Hz, while Experiment 2 repeated the experiment at a center frequency of 8 kHz. If a sampling mechanism uses temporal cues to extract the pitch of successive samples, then the results at 8 kHz should be different than those at 500 Hz.

2. Experiment 1

2.1 Materials and methods

2.1.1 Subjects

7 adult listeners participated in the study (2 female, M = 26.14 years, SD = 4.91 years), all with normal hearing and no tonal language experience. Participants were financially compensated for their time and all procedures were approved by the University of California, Irvine Institutional Review Board.
2.1.2 Stimuli and equipment

Stimuli were unidirectional logarithmic frequency sweeps, generated using the following equation (Hsieh & Saberi, 2009):

\[
Y(t) = \sin \left( 2\pi \frac{f_s}{\ln(f_e/f_s)} \left[ \exp \left( \frac{\ln(f_e/f_s)}{T_s} t \right) - 1 \right] \right)
\]

[3]

where \( T_s \) is the stimulus duration and \( f_s \) and \( f_e \) are the start and end frequencies of the sweep, in hertz.

On each trial, the stimulus parameters of rate and transition span were independently sampled from uniform distributions; rate varied between 0.5 semitones per second (which equates to 0.042 octaves/sec) and 2 semitones per second (0.167 octaves/second), and transition span varied between 0.1 and 0.5 semitones. Stimulus duration was not directly manipulated but was instead derived from stimulus rate and transition span, and varied between 50 ms and 1000 milliseconds.

All sweeps had a nominal center frequency of 500 Hz. To prevent listeners from using absolute frequency cues to determine sweep direction, the center frequency of each sweep was uniformly roved in a 4 semitone range around a center frequency of 500 Hz; all stimuli were thus between 396 and 630 Hz. Sweeps moved either up or down, and the order of sweeps was randomized across trials. All sweeps had 10 ms on/off ramps and were played over Sennheiser HD 380 pro headphones at a sampling rate of 44.1 kHz at 70 dB SPL.

On each trial, listeners were asked to judge whether a sweep moved up or down, and indicated their choice via keypress. Listeners performed a total 6000 trials in three separate 1-hour sessions (2000 trials per session). Visual feedback was provided.
2.1.3 Data analysis

To analyze the data, stimulus parameters of rate and transition span were binned into 7 equal-sized bins, with center values shown in Table 2.1. For each subject, the responses within each bin were pooled and sensitivity index ($d'$) was calculated (Green & Swets, 1966), yielding 49 total $d'$ values for each subject. $d'$ was then averaged across all 7 subjects.

**TABLE 2.1.** Center values of the 7 bins for stimulus parameters of rate and transition span, expressed in semitones per second (for rate) and semitones (for transition span).

<table>
<thead>
<tr>
<th>Rate (semitones/sec)</th>
<th>Transition Span (semitones)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6071</td>
<td>0.1288</td>
</tr>
<tr>
<td>0.8214</td>
<td>0.1857</td>
</tr>
<tr>
<td>1.0357</td>
<td>0.2429</td>
</tr>
<tr>
<td>1.2500</td>
<td>0.3000</td>
</tr>
<tr>
<td>1.4643</td>
<td>0.3571</td>
</tr>
<tr>
<td>1.6786</td>
<td>0.4143</td>
</tr>
<tr>
<td>1.8929</td>
<td>0.4714</td>
</tr>
</tbody>
</table>

2.2 Results

2.2.1 Psychophysical results

Mean direction identification sensitivity ($d'$) of all subjects is shown as a colormap in Figure 2.1. Stimulus parameters of rate and transition span are displayed on the x-axis and y-axis, respectively, and color indicates direction sensitivity. A 2-way repeated measures ANOVA with fixed-effects variables of rate and transition span revealed a significant main
effect of both rate \( F(1.8, 11.1) = 19.06, p < 0.001 \) and transition span \( F(1.2, 7.1) = 28.52, p < 0.001 \), and a nearly significant interaction between the two, \( F(4.2, 25.1) = 2.64, p = 0.055 \). In other words, both rate and transition span had a significant effect on direction identification sensitivity. Increasing the rate of frequency sweep led to poorer sensitivity and increasing the sweep transition span led to better sensitivity. Thus, sweeps with rapid changes in frequency were more difficult to identify.

**Figure 2.1.** Colormap of direction identification sensitivity (d’) as a function of sweep rate and transition span for sweeps with nominal center frequencies of 500 Hz. Color represents d’, with the lowest d’ values in the lower right-hand portion of the graph, and the highest d’ values in the upper left corner. Maximum d’ was 3.3. Black lines overlaying the colormap represent iso-duration lines (iso-duration lines do not visually appear straight due to rounding errors in MATLAB). Black numbers indicate approximate stimulus duration in milliseconds.
Although previous studies have not directly evaluated the effect of sweep rate in the perception of sweep direction, these data are analogous to previous studies of sweep perception that find increased detection and discrimination thresholds when rate of frequency change increases (Thyer & Mahar, 2006)

2.2.2 Analytical modeling

Figure 2.2. Predicted direction identification sensitivity (d’) as a function of sweep rate and transition span for sweeps at 500 Hz. Colored plane represents predicted d’ values. Black dots indicate behavioral d’ data at each of the 49 points. Opaque dots represent performance that falls below the predicted performance plane. The maximum predicted d’ was 4.19 and the minimum was –0.048. The limits of the color scale are –0.05 to 4.19.
Linear regression was used to predict direction identification performance, and showed that $d'$ can be accurately estimated from a weighted linear combination of transition span and rate, with the equation:

$$d' = 0.357 + 9.35t - 889.27r, \quad R^2 = .97$$

Where $t$ indicates transition span and $r$ indicates rate.

2.3 Discussion

The fact that sensitivity decreased as rate of frequency change increased – i.e. that sweeps with faster rates of frequency change were more difficult to identify – provides evidence for a sampling mechanism in the perception of sweep direction: as rate of change increases (and duration decreases), the estimates of instantaneous frequency at each sample becomes less precise and results in poorer performance.

These data also provide preliminary support for the temporal coding model, because phase-locking cues are also adversely affected by rapid rates of change. However, Experiment 1 does not provide incontrovertible evidence for phase-locking cues because a sampling mechanism based on energy cues may also be disrupted by decreased duration, although to a lesser extent. Experiment 2 aims to clarify the role of phase-locking cues for this task.

3. Experiment 2

Experiment 2 thus replicates Experiment 1 but at higher frequencies where phase-locking is unreliable. If phase-locking cues are contributing to the estimate of start/end pitch cues and the perception of sweep direction at low frequencies, then performance at
these higher frequencies should differ from performance at low frequencies; it should either be poorer if no other cues are available at high frequencies, or should be less adversely affected by rate if energy cues are used at high frequencies. This differential effect of rate would arise from the sluggish nature of phase-locking cues at low frequencies (Moore & Sek, 1995, 1996; Sek & Moore, 1995).

3.1 Materials and Methods

3.1.1. Subjects

Seven adult listeners participated in the experiment (2 females, M = 29.28 years of age, SD = 1.68 years of age), 5 of whom also took part in Experiment 1. All participants had normal hearing, none spoke a tonal language, and none had absolute pitch.

3.1.2. Stimuli and procedure

The stimuli and procedure of Experiment 2 were identical to Experiment 1 except that all stimuli were roved in a 2-semitone range around 8 kHz. The choice of a 2-semitone uniform range rather than a 4-semitone range was to ensure that the stimuli were neither too high to be uncomfortable to subjects, nor too low that they fell into a frequency range where phase-locking was available. All stimuli were between 7,127 and 8,980 Hz.

Stimuli were again presented over Sennheiser HD 380 pro headphones. To eliminate any level cues that might be introduced by the headphones and ear canal resonances at high frequencies and ensure that all stimuli were presented at the same sound level, the headphone’s transfer function was mapped using a Brüel & Kjær, Model 2260 precision sound level meter with a 6-cc coupler and 0.5-in. microphone (Brüel & Kjær, Model 4189),
and all stimuli were inverse filtered to guarantee an at-ear sound level of 70 dB SPL. At the low frequency range, the headphone transfer function is relatively flat and therefore does not require correction as much as at higher frequencies.

3.2. Results

3.2.1. Psychophysical results

Figure 2.2 shows $d'$ as a function of rate and transition span, averaged across all subjects. A 2-way ANOVA with fixed-effects variables of rate and transition span revealed a significant main effect of rate [$F(3, 18.1) = 6.605, p = 0.003$] and of transition span [$F(1.18, 7.07) = 31.15, p = 0.001$]. i.e., similar to experiment 1, increasing the transition span led to better identification sensitivity, and increasing the rate of frequency change led to poorer direction identification sensitivity. There was also a significant interaction between rate and transition span, $F($

To compare across frequency regions, a 3-way ANOVA with fixed-effects variables of rate and transition span and random-effects variable of frequency region was performed on the data from both experiment 1 and 2, and again revealed a significant main effect of rate [$F(2.4, 28.5) = 24.6, p < 0.001$] and a significant main effect of transition span [$F(1.2, 14.5) = 58.67, p < 0.001$]. Although there was no significant main effect of frequency region [$F(1, 12) = 0.212, p = 0.654$], indicating that $d'$ values did not significantly differ between Experiments 1 and 2, there was a significant interaction between rate and frequency region, $F(2.4, 28.5) = 4.3, p = 0.018$, indicating that rate had a greater effect on $d'$ in the low frequency region than at high frequencies.
Figure 2.3. Colormap of direction identification sensitivity (d’) as a function of sweep rate and transition span for sweeps with nominal center frequencies of 8 kHz. Color represents d’, with the lowest d’ values in the lower right-hand portion of the graph, and the highest d’ values in the upper left corner. Maximum d’ was 3.3. Black lines overlaying the colormap represent iso-duration lines (iso-duration lines do not visually appear straight due to rounding errors in MATLAB). Black numbers indicate approximate stimulus duration in milliseconds.

3.2.2 Analytical modeling

Linear regression was again used to predict direction identification performance, and showed that d’ can be accurately estimated from a weighted linear combination of transition span and rate, with the equation:

\[
d’ = -0.122 + 7.97t - 377.21r, \quad R^2 = .97
\]
Where \( t \) indicates transition span and \( r \) indicates rate. Figure 2.4 shows the predicted \( d' \) values.

**Figure 2.4.** Predicted direction identification sensitivity (\( d' \)) as a function of sweep rate and transition span for sweeps with nominal center frequencies of 8 kHz. Colored plane represents predicted \( d' \) values and black dots indicate behavioral \( d' \) data at each of the 49 points. Opaque dots represent performance that falls below the predicted performance plane. The maximum predicted \( d' \) was 3.38 and the minimum was 0.22. Limits of the color scale are from 0.22 to 3.38.
3.3. Discussion

Similar to experiment 1, experiment 2 found that both the rate and transition span of sweeps affected how well subjects were able to identify whether sweeps moved up or down; subjects were better able to identify sweep direction when the sweeps had larger transition spans and when they moved more slowly through frequency space. Like Experiment 1, Experiment 2 again provides evidence for a sampling mechanism. A comparison of results between experiments 1 and 2 show that, although there was no significant difference in the overall magnitude of $d'$ between frequency regions, increasing the sweep’s rate of frequency change was more detrimental to performance at low frequencies than at high frequencies. This suggests that a sampling mechanism extracts information about start/end pitches using temporal information and phase-locking at low frequencies, but uses a different mechanism at high frequencies. Because performance at high frequencies is still adversely affected by rate, performance at high frequencies may still use a sampling mechanism to determine the direction of pitch movement, just one based on rate-place coding instead of temporal processing.

4. General Discussion

The current study evaluated the mechanisms underlying the perception of direction of FM sweeps, specifically the use of a sampling mechanism and phase-locking cues in the perception of FM sweep direction. To our knowledge, this is the first direct investigation of these cues in direction perception; the few previous studies investigating the perception of sweep direction found evidence for a mechanism that operated in addition to a dynamic mechanism but did not specifically examine whether this alternate mechanism was a
sampling mechanism based on temporal pitch processing (Demany et al., 2009; Tsumura et al., 1973). The studies presented here used sweeps with slow rates of change and varied parameters of rate and transition over a wider range than many previous studies have used, allowing for a better characterization of how performance is affected by stimulus parameters of rate and transition span. To directly address the question of whether phase-locking is used in a sampling mechanism, we also included a frequency region manipulation.

In both frequency regions, direction identification sensitivity worsened with increasing sweep rate. This is not a result that one might expect if sweeps were identified via FM-direction-selective cells, since these cells are not adversely affected by rate, and in fact may function better for faster rates (Demany et al., 2009; C. Ye, Poo, Dan, & Zhang, 2010; Zhang et al., 2003). Thus, these data support a mechanism that is different from the dynamic mechanism, which may be the sampling-differencing mechanism based on Hartmann and Klein’s 1980 model, in which listeners extract the pitches at successive time points (or at the start/end) of the sweeps and compare them to determine sweep direction. Previous studies of FM and sweep perception have suggested that such a mechanism may be useful for the perception of slower sweeps (Demany et al., 2009). Because the sweeps in the current experiments are as slow as 1/24 an octave per second, and thus much slower than those used in the study of FM-selective cells, this remains possible.

In examining the cues used to form snapshots, these data provide support for the use of phase-locking cues at low frequencies. Although we did not see a sharp drop in performance between experiment 1 and experiment 2, which would be a strong indication of the use of phase-locking cues at low frequencies, we did find that increasing rate
(decreasing duration) was less disruptive at high frequencies than at low frequencies. Because a sampling mechanism based on spectral cues is less adversely affected by rapid rates than a sampling mechanism based on phase-locking cues, these data suggest that a sampling mechanism is used at both low and high frequencies, but is reliant on phase-locking cues at low frequencies and not at high frequencies.

The role of a sampling mechanism in the perception of FM and frequency sweeps is ubiquitous, but is found predominantly in studies that address the use of this mechanism for the detection and discrimination of sweeps, and less often in studies of the perception of the direction of sweeps. The series of experiments reported here address this discrepancy and provide support for the use of a sampling mechanism in direction identification for slow frequency sweeps. Although such a sampling mechanisms is commonly assumed to involved phase-locking cues, the current study directly investigated this assumption, and confirmed that phase-locking is likely used in a sampling mechanism at low frequencies. The data furthermore suggest a difference in listening strategies between the two frequency regions, although a sampling mechanism is likely used in both regions.

This study has several limitations that should be considered in a discussion of the results. First, most of the subjects who participated in Experiment 2 had previously participated in Experiment 1, introducing the possibility of training effects. Although a comparison of naïve versus experience subjects did not reveal a significant difference in direction identification sensitivity, there were only 2 naïve subjects and there could be insufficient statistical power to see any difference. Second, it is possible that ear canal resonances, introduced by the circumaural headphones added spectral or loudness cues
that provided an additional sweep direction cue that allowed subjects to perform better in experiment 2. These ear canal resonances are prevalent in the high frequency regions, and while the headphones were placed carefully over subjects’ ears and not moved during the experiment, and all frequencies were inverse filtered to eliminate headphone-generated level variations, it is possible that some cues were present. Regardless, both training effects and confounding spectral cues would have led to improved performance at 8 kHz relative to 500 Hz, which would have obscured any effects of phase-locking rather than enhancing them.
CHAPTER 3:

Standard interval size affects interval-discrimination thresholds

Pitch changes are also the foundation of musical composition, where the distance by which a pitch changes – defined as a *pitch interval* or *musical interval* – has special relevance. Musical intervals, quantified in semitone units, form melodies when they are combined sequentially and harmonies when they are combined simultaneously. In Western musical theory, different intervals serve different functional roles and convey different emotions. The interval of 7 semitones, called a perfect fifth, is used to create harmonious and consonant sounds, while the 6-semitone interval, called an augmented 6th or *tri-tone*, was historically used to create musical tension and dissonance (Cooke, 1959). Pitch-interval perception and our ability to discriminate between pitch intervals of different sizes is thus an important aspect of both speech and music perception, and is commonly studied in auditory perceptual research.

Such studies of pitch-interval perception use a variety of paradigms ranging from the method of adjustment (Demany & Semal, 1990; Plomp & Steeneken, 1968; Ward, 1954) and subjective ratings (Kameoka & Kuriyagawa, 1969; McDermott, Lehr, & Oxenham, 2010; Plomp & Levelt, 1965; F. A. Russo & Thompson, 2005; van de Geer, Levelt, & Plomp, 1962), to interval identification and discrimination (Burns & Campbell, 1994; Burns & Ward, 1978; Killam, Lorton, & Schubert, 1975; Siegel & Siegel, 1977a; Zatorre & Halpern, 1979). Interval identification paradigms, which require listeners to name intervals of the Western musical system using labels such as “minor second” and “perfect fifth”, have shown that musicians demonstrate learned categorical perception for the 12 canonical intervals of the
Western musical system (e.g. 1 semitones or 2 semitones) (Burns & Campbell, 1994; Burns & Ward, 1978; Siegel & Siegel, 1977a, 1977b; Zatorre & Halpern, 1979) and have greater difficulty identifying non-canonical intervals such as quarter tones (i.e. 2.5 semitones or 0.5 semitones) without explicit training (Siegel & Siegel, 1977b). Musical training also enhances interval perception by enabling listeners to more easily detect changes made to a single note embedded in a short musical melody (Dowling, 1978; Dowling & Fujitani, 1971); listeners without musical experience can only detect brief melodies with altered intervals if the contour of these melodies is also altered. Additional studies of interval perception show that relative pitch perception is influenced by a variety of other stimulus factors, including harmonicity (McDermott, Keebler, Micheyl, & Oxenham, 2010; Plomp, Wanegaar, & Mimpen, 1973; Trainor, 1996), timbre (F. A. Russo & Thompson, 2005; Zarate, Ritson, & Poeppel, 2013), sound level (Thompson, Peter, Olsen, & Stevens, 2012), and whether the interval is ascending/descending or simultaneous/sequential (Killam et al., 1975; Luo et al., 2014; for a review, see Burns, 1999).

Due to the highly musical nature of interval perception, many paradigms feature stimuli in a musical context (as in the short melodies above), or require a minimal amount of musical experience by the listeners (as in the musical interval identification tasks). But because this often precludes the possibility of examining this perceptual ability in listeners without formal musical training, many investigators instead use pitch-interval discrimination tasks to avoid reliance on musical experience. In a pitch-interval discrimination task, also called an *interval discrimination task*, listeners are presented with two intervals and are asked to judge which is larger. This task is similar to basic frequency discrimination tasks except that listeners are asked to identify the larger interval rather
than the higher tone. Not surprisingly, such studies show that discrimination performance improves as the difference between two intervals increases and that musicians typically perform the task better than non-musicians, even without an explicit musical context (Burns & Ward, 1978; Luo et al., 2014; McDermott, Keebler, et al., 2010; Zarate, Ritson, & Poeppel, 2012; Zarate et al., 2013).

Yet, studies of pitch-interval discrimination report conflicting results about how listeners are affected by the size of the reference interval. Listeners in several studies (Burns & Ward, 1978; McDermott, Keebler, et al., 2010) produced thresholds which did not significantly differ across different standard intervals. The minimum difference (quantified in semitones units) needed to discriminate a 1-semitone interval from a slightly larger interval is the same as needed to discriminate between a 4-semitone interval and a slightly larger one. This trend holds true for standards that are both canonical (1-semitones, 2-semitones) and non-canonical (1.5-semitones, 2.5-semitones) Western musical theory intervals (McDermott, Keebler, et al., 2010). However, a more recent study examining a wide range of standard interval sizes showed that discrimination thresholds strongly varied with standard interval size and were higher for larger standard intervals, increasing by an average of 0.22 semitones for each interval-standard increase of 1-semitone (Luo et al., 2014).

The different effects of standard-interval size across studies may stem from a number of factors, including the extent of the base tones' frequency rove, whether the stimuli are pure or complex tones, and, in particular, the musical experience of the listeners. The existence of an effect of standard interval magnitude for certain listeners may have important implications for our understanding of relative pitch perception.
Furthermore, if the effect is influenced by musical training, an examination of this type of perception may shed light on potential differences between the listening strategies of musicians and nonmusicians.

The goal of the current study was thus twofold: to examine how listeners’ pitch-interval discrimination thresholds vary with the size of the standard interval across large intervals, and to examine whether previous conflicting reports of the effects of standard interval size might be due to differences in musical training. To this end, listeners were tested in a melodic pure-tone interval-discrimination task with a procedure analogous to that used by McDermott et al. (2010a) and Luo et al. (2014) to explore pitch-interval discrimination by nonmusicians, amateur musicians, and professionally trained musicians across three standard-interval conditions: 1, 6, and 7 semitones. Standard intervals of 1, 6, and 7 semitones are sufficiently large so as to show a possible interval-magnitude effect, but not so large that they introduce problems with frequency roves. We included both 6- and 7-semitone conditions because, although similar in semitone size, these two intervals are radically different in their functional and theoretical role in Western musical theory. They are also larger than many standards used previously. If an effect of standard-interval magnitude exists for the melodic intervals used in the current study, then thresholds for the 6- or 7-semitone conditions should be significantly larger than those obtained using a 1-semitone standard.

Since musical training is well-known to affect pitch and interval perception (Kishon-Rabin, Amir, Vexler, & Zaltz, 2001; Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Spiegel & Watson, 1984), and differences in the musical experience of the listeners may explain previous discrepancies in the effect of interval size, both musicians and
nonmusicians were tested. Furthermore, because the degree of musical training can highly vary from musician to musician, we further separated the musicians into subgroups: those who had formal music theory instruction which included one year of ear training and those who did not receive such formal training. Formal music theory instruction includes training in the harmonies, tonalities, and intervals of the Western musical system, and when taught in a university setting is almost always paired with a standardized ear training/sight-singing curriculum. Ear training/sight-singing classes teach students how to recognize intervals, discriminate between them, and vocally produce them without reference tones. These tasks are practiced either outside of a musical context or with the intervals embedded in a musical melody, and are intended to develop a musician’s sense of relative pitch. Since ear training/sight-singing is often a required part of an undergraduate or graduate degree in music, all participants with formal music theory instruction and ear training were either currently pursuing or had already earned a bachelor’s degree in music or as a music minor. All members of this group also had at least 12 years of musical experience. It was therefore expected that although musicians would perform better than nonmusicians in general, ear training’s emphasis on interval recognition and identification may allow formally trained musicians to show additional improvements in discrimination performance and produce thresholds that would be consistent across different interval standards, potentially clarifying our understanding of the differences between the interval-discrimination results of previous studies.
2. Methods

2.1. Listeners

Fourteen adult listeners participated in the study. All reported normal hearing, none had absolute pitch, and none spoke a tonal language. Listeners were classified into three groups based on musical experience. Five nonmusicians (all males, mean (M) = 25.6 years of age, standard deviation (SD) = 3.7 years) had less than three years of musical instruction during childhood. Three of the nonmusicians had no musical experience and two had 9 months and 3 years respectively, both at least ten years prior to the experiment (M = 0.75 years of experience, SD = 1.3 years of musical experience). The nonmusician with 3 years of musical training reported that this training was intermittent and informal, and as a result he could not play an instrument or read music. Three amateur musicians (1 male, 2 females, M = 23.7 years of age, SD = 2.9 years) had between 10 and 12 years of music lessons but no formal music theory instruction (M = 10.7 years of musical experience, SD = 1.2 years of musical experience). All amateur musicians reported that they regularly played music recreationally, and were considered amateur because they had never studied music theory or received formal ear training. Six expert musicians (4 males and 2 females, average age M = 26.2 years, SD = 5.6 years) were either currently pursuing or had earned a Bachelor’s degree as either a music major or music minor, had at least 12 years of music lessons, and 1 year of a college ear training/sight-singing course (M = 16.8 years of experience, SD = 6.7 years). All expert musicians were actively engaged in recreational musical composition and practice. Of the six expert musicians, two were classically trained in both piano and violin, two were classically trained in violin, piano and guitar, one was a jazz violin and mandolin player, and one was a jazz drummer. Participants were financially
compensated for their time and all procedures were approved by the University of California, Irvine Institutional Review Board.

2.2. Stimuli and equipment

Each trial contained four sequential pure tones grouped into two melodic pitch intervals (Figure 3.1). Tones 1 and 2 defined interval A and tones 3 and 4 defined interval B. All tones were 500 ms in duration with 20 ms linear on/off ramps. The two tones of each interval were separated by a 250-ms gap and the two intervals of each trial were separated by a 1000-ms gap. All stimuli were generated in Matlab (Natick, MA) and played over Sennheiser HD 380 pro headphones at a sampling rate of 44.1 kHz. To ensure that every listener heard all tones of the experiment at the same sound level, stimuli were inverse filtered with the headphone transfer function to generate an at-ear-drum sound

Figure 3.1. Sample stimuli for a single trial. Tones 1 and 2 form Interval A and tones 3 and 4 form Interval B. The red line corresponds to size of smaller interval and the blue line represents the size of the larger interval. The frequencies of tones 1 and 3 are randomized across trials.
level of 70 dB SPL for all tones, measured with a 6-cc coupler, 0.5-in. microphone (Brüel & Kjær, Model 4189) together with a Brüel & Kjær Model 2260 precision sound level meter. All testing was administered in a steel double-walled acoustic sound booth (Industrial Acoustics Company). Participants indicated their choice using a mouse click and visual feedback was provided after each trial.

All intervals were ascending melodic intervals, in that the second tone of each interval was always higher in frequency than the first. The two intervals (A and B) of a trial were always unequal in magnitude, and listeners were instructed to indicate which was larger. The sizes of the two intervals of each trial are designated \( i \) and \( i + \Delta i \), where \( i \) is the magnitude of the smaller standard interval (fixed within a run) and \( i + \Delta i \) is the magnitude of the comparison interval. \( \Delta i \) is the difference in magnitude between the two intervals and its value for each trial was determined via an adaptive staircase method (see section 2.3). The order of presentation of the two intervals was counterbalanced across trials.

To ensure that listeners were not able to perform the task by comparison of tones 2 and 4, tones 1 and 3 were roved over a continuous range of ±9 semitones around a center frequency of 220 Hz and were always at least 4 semitones apart. Monte Carlo simulations confirmed that these roves were large enough that the task could not be adequately performed by choosing the interval with the highest tone.

2.3. Adaptive procedure

The value of \( \Delta i \) was controlled by a 2-down-1-up adaptive procedure that converged on 70.7% correct (Levitt, 1971). At the start of each run, \( \Delta i \) was set to 12 semitones. It was decreased after two consecutive correct trials, and increased after 1 incorrect trial. At each
step, $\Delta i$ was either divided by a factor of $10^y$ ($\Delta i$ following two consecutive correct trials) or multiplied by a factor of $10^y$ (following an incorrect trial), where $y = 0.2$ up to and including the $4^{\text{th}}$ reversal, and $y = 0.05$ thereafter.

Runs ended after the 50th trial. At the conclusion of each run, the threshold (in semitones) was calculated from the geometric mean of the values of $\Delta i$ at all reversals except the first 3 (for runs with an odd number of reversals) or the first 4 (for runs with an even number of reversals). The mean number of reversals per run was 12.3. The minimum number of reversals was 5 and the maximum number of reversals was 21 (both of which occurred once).

2.4. Experimental design and procedure

The experiment followed a $3 \times 3$ mixed design (3 standard intervals by 3 levels of musical experience). Each participant performed 5 runs per condition, and all 5 runs were completed before moving to the next condition. The order of conditions was pseudo-randomized across participants such that each condition was performed first at least once, and last at least once.

Data were collected in two sessions conducted on consecutive days. Participants completed either 1 or 2 conditions per day. Each session began with a 10-minute training period designed to familiarize participants with the task and interface. The training period consisted of a sample run of 50 trials in which all adaptive parameters were identical to those for experimental runs, except that the value of $i$ was 2 semitones and the frequencies of the tones were roved in a continuous ±9 semitone range around 440 Hz (rather than 220 Hz); this was done to ensure that the training procedure was identical to the experimental
procedure, but the stimuli were different enough to avoid biasing the results of any given condition. Following training, participants completed all runs of the session. Participants were given a minimum 3-minute break after finishing each condition, and longer breaks if requested, but were asked to complete all runs of a condition before pausing.

3. Results

3.1 Effects of musical experience

Figure 3.2 shows the mean pitch-interval discrimination thresholds for the three subject groups averaged across the three stimulus conditions (1, 6, 7 semitones). A between-subjects ANOVA revealed a significant effect of group \( F(2,11) = 6.238, p = 0.015 \) with expert musicians producing the lowest thresholds of the three groups, in agreement with previous studies demonstrating that musicians perform better than nonmusicians on

![Figure 3.2](image)

**Figure 3.2.** Mean thresholds for three subject groups, averaged across all interval-standards. Expert musicians produced the lowest thresholds of the three groups. Error bars represent one standard error of the mean.
pitch- and interval-discrimination tasks (Kishon-Rabin et al., 2001; McDermott, Keebler, et al., 2010; Micheyl et al., 2006; Spiegel & Watson, 1984; Zarate et al., 2012, 2013).

Independent t-tests reveal showed a significant difference between the thresholds for the non-musicians and the expert musicians \([t(9) = 2.90, p = 0.018]\) and between the amateur musicians and the expert musicians \([t(6.360) = 3.74, p = 0.009]\), but no significant difference between the thresholds for the amateur and non-musicians \([t(5.163) = 0.088, p = 0.933]\). Thus, although the results for the expert musicians are consistent with previous studies showing differences between musicians and nonmusicians in pitch and interval-perception tasks (Kishon-Rabin et al., 2001; McDermott, Keebler, et al., 2010; Micheyl et al., 2006; Spiegel & Watson, 1984; Zarate et al., 2012, 2013), but the results for the amateur musicians who performed similarly to nonmusicians are not consistent with previous findings.

A one-way ANOVA was performed on the mean thresholds for the 1-semitone condition. This also showed a significant difference between the thresholds of the three subject groups, \(F(2, 11) = 9.594, p = 0.004\). Independent t-tests again revealed a significant difference between the non-musicians and the expert musicians \([t(9) = 3.91, p = 0.004]\) and between the amateur musicians and the expert musicians \([t(7) = 3.40, p = 0.011]\), but not between the amateur and non-musicians \([t(6) = 0.857, p = 0.424]\).
3.2. Effects of standard-interval size

Figure 3.3 shows the threshold for each group in each condition. The error bars for amateur musicians and nonmusicians are offset slightly to the left of each data point to facilitate visual comparison. Thresholds were lowest in the 1-semitone condition for all listeners, and were notably lower than the thresholds in the 6- and 7-semitone conditions (Figure 3.2). A 3 x 3 mixed analysis of variance showed a significant main effect of standard interval magnitude \[ F(2,22) = 16.425, p < 0.001 \], and no significant interaction between standard interval size and subject group \[ F(4,22) = 0.77, p = 0.556 \]. A contrast analysis revealed a significant difference between the thresholds for the 1-semitone standard condition and the 6- and 7-semitone conditions \[ F(1,11) = 17.618, p = 0.001 \], but no significant difference between the 6- and 7-semitone conditions, \[ F(1,11) < 0.001, p = 0.991 \].

**Figure 3.3.** Mean thresholds for the three subject groups and for standard interval distances of 1, 6, and 7 semitones. Thresholds are lower for a 1-semitone standard interval than for 6- or 7-semitone standard interval magnitudes. Expert musicians (green) produced the lowest thresholds in all conditions. Error bars represent 1 standard error of the mean. To facilitate visual inspection, error bars for nonmusicians (blue) and amateur musicians (red) are offset slightly to the right and left of the data points, respectively.
The increase in thresholds with increasing standard-interval size markedly differs from the results of Burns and Ward (1978) and McDermott et al. (2010a) but is consistent with the results of Luo et al. (2014).

To test whether the formal music theory experience of our expert musicians affected their performance, and to compare the results with those of previous studies that focused on the performance of highly trained musicians in relative pitch tasks (Burns & Campbell, 1994; Burns & Ward, 1978), planned contrasts were performed on the thresholds for the expert musician group. These showed a nonsignificant effect of standard interval size, F(2,10) = 3.770, p = 0.060.

### 3.3 Thresholds expressed as a proportion of standard interval size

To compare thresholds across subject groups as a proportion of standard magnitude, thresholds were expressed as Weber fractions by dividing the threshold Δi by i (Table 3.1 and Figure 3.4). A 2-way ANOVA performed on these fractions revealed a significant main effect of group [F(2,11) = 13.43, p = 0.001], with expert musicians exhibiting lower fractions than amateur musicians and nonmusicians.

**Table 3.1.** Thresholds expressed as a Weber Fractions (Δi/i)

<table>
<thead>
<tr>
<th></th>
<th>1-semitone</th>
<th>6-semitones</th>
<th>7-semitones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonmusicians</td>
<td>2.05</td>
<td>0.74</td>
<td>0.68</td>
</tr>
<tr>
<td>Amateur musicians</td>
<td>1.66</td>
<td>0.77</td>
<td>0.69</td>
</tr>
<tr>
<td>Expert musicians</td>
<td>0.58</td>
<td>0.39</td>
<td>0.27</td>
</tr>
</tbody>
</table>
There was also a significant effect of standard interval size \[F(1.14,12.53) = 20.575, p < 0.001\] with fractions decreasing as standard interval size increased, and a significant interaction effect \[F(2.28,12.527) = 3.773, p = 0.048\].

\[\text{Figure 3.4.} \quad \text{Pitch-interval discrimination thresholds expressed as Weber fractions (}\Delta i/i)\text{ for each subject group as a function of standard interval size. Fractions decrease as standard interval increases and expert musicians produced the lowest fractions of the three groups. Error bars omitted for clarity.}\]

\[4. \text{Discussion}\]

These results show that, in an interval-discrimination task using pure tones, thresholds vary across different standard interval magnitudes. For the standard intervals tested in this study, discrimination thresholds were very high for standard interval sizes of 6- and 7-semitones. Intervals of this size have not previously been used in a direct comparison of musician and non-musician listeners, and the data for these larger intervals clearly demonstrates that intervals of 6 and 7 semitones exhibit different thresholds than
intervals of 1 semitones. In agreement with previous studies, we found that thresholds for expert musicians were significantly lower than those for our other listeners, while the amateur musicians produced thresholds that were the same as for the nonmusicians.

The effect of standard interval size is consistent with the results of Luo et al. (2014). On the other hand, the results differ from those of multiple previous studies in which the thresholds were consistent across conditions with different standard-interval sizes (Burns & Ward, 1978; McDermott, Keebler, et al., 2010). However, there were notable differences across studies: while Luo et al. (2014) tested a wide range of interval sizes that may have accentuated an effect of standard interval size, they did not specifically categorize listeners according to musical experience. And while Burns and Ward (1978) and McDermott et al. (2010a) included musicians and nonmusicians, they only used intervals smaller than 6 semitones; despite this limited range, the thresholds found by McDermott et al. (2010a) showed a slightly increasing linear trend across standard interval sizes. The current study tested nonmusicians, informally trained musicians, and formally trained musicians with intervals as large as 7 semitones, and found a significant difference between thresholds from small and large intervals. Although we expected expert musicians to show a different effect of standard interval size than nonmusicians and amateur musicians due to their formal music theory instruction and ear/training sight-singing experience, the lack of a significant interaction between musicianship and standard size does not support this expectation.

The fact that discrimination thresholds increased as the size of the standard interval increased might reflect Weber’s Law, which states that the just-noticeable difference increases as the magnitude of the stimulus increases. However, when thresholds were
expressed as a proportion of standard magnitude, i.e. as $\Delta i/i$, they were not constant across conditions. Rather, the $\Delta i/i$ ratio decreased as the magnitude of the standard interval increased. In other words, while thresholds increased across the particular standard magnitude conditions we studied, they did not do so as a constant proportion of standard magnitude, and thus show a deviation from Weber’s law.

Small intervals are more common in both speech and music (Vos & Troost, 1989), and listeners’ greater familiarity with them may have exacerbated the differences between the thresholds for the 1-semitone interval and the 6- and 7-semitone intervals.

Only the expert musicians produced thresholds that were significantly different from those for the nonmusicians, while the amateur musicians gave results nearly identical to those for the nonmusicians. This suggests that there may be something about the expert musicians in this study that differentiates them from amateur musicians. The primary difference between the two musician groups is the nature of their musical experience: the expert musicians have formal university music theory training and at least one year of ear training/sight-singing, while amateur musicians only had music lessons. Although the expert musicians also had more years of musical experience than the amateur musicians (an average of 16.8 years for the expert musicians and 10.7 years for the amateur musicians), the difference between the amateurs and experts was less than that between amateurs and nonmusicians. There may be something about formal music theory and ear training of the expert musicians that affects pitch-interval discrimination thresholds in a way that music lessons or time spent playing an instrument does not. This has interesting implications for music theory instruction and auditory training: while musical training is known to enhance perception in a variety of ways, these benefits often require several
years of musical experience. But if ear training is indeed effectively training relative pitch listening skills, then this may be one way that the benefits of structured auditory training – delivered over the course of 9-months to a year rather than the multiple years required of typical musical training – can generalize to non-musical environments. However, because we were unable to test pitch-interval discrimination before and after musical training occurred, and because both Western music theory’s ear training and our current task specifically target the intervals of the Western musical system, further studies are needed to investigate how ear training/sight singing might benefit listeners in a variety of non-musical auditory tasks that extend beyond interval perception.

When considering the relative pitch and interval processing abilities that would be required for music perception in general and that have been documented in prior studies, the discrimination performance measured here may seem surprisingly poor. The high thresholds found here, especially for listeners without formal musical training, may have resulted from several factors. First, the stimuli were pure tones and thus lacked many of the additional qualities – such as timbre and harmonicity – of the stimuli generated by natural musical instruments. Secondly, the current procedure included a frequency rove of the initial tones of each interval, which would have increased stimulus uncertainty and made the task more difficult. Lastly, listeners were asked to make judgments about the relationship between the two intervals of each trial (e.g. “which interval is wider?”) rather than simply detect a change in repeating stimuli (see Schellenberg and Trainor, 1996; Schellenberg and Trehub, 1994; Trainor, 1997). This is a task that would have been particularly challenging to listeners since humans may have poor interval-perception in general (McDermott, Keebler, et al., 2010), although it is worth noting that even changes
made to intervals within short musical melodies can be difficult for some listeners to detect if those changes do not violate the overlearned harmonic structure of a melody or its contour (Dowling, 1978). Although the stimuli and task of the current study presented a challenge for participants, these findings nevertheless extend the results of previous pitch-interval discrimination studies and highlight the effect of standard interval size.

5. Conclusions

Pitch-interval discrimination thresholds were found to vary with the size of the standard interval. In other words, when listeners are discriminating between a pitch interval of size $i$ and one slightly larger ($i + \Delta i$), the minimum threshold ($\Delta i$) needed to discriminate the larger one depends on the size of the standard interval $i$. The results suggest a potential added benefit of formal music theory and ear-training which is different from that of simple musical experience: musicians with formal musical training produced the lowest thresholds, while musicians without this training performed similarly to nonmusicians.
CHAPTER 4:

Effects of frequency region on interval discrimination thresholds

Basic pitch perception is inherent to pitch-interval perception, yet there are important differences between the two. In basic pitch perception we perceive and make judgments about the exact pitch of an isolated tone, while in interval perception we make judgments about the frequency relationship between two tones regardless of their absolute frequencies. Although we have very high perceptual acuity for basic pitch perception, measured via frequency difference limens that can be as low as 10 Hz (Plack & Oxenham, 2005), researchers have found that these benefits do not extend to interval perception (McDermott, Keebler, et al., 2010).

Researchers have also documented stimulus manipulations that influence interval perception but leave basic pitch perception unaffected (F. A. Russo & Thompson, 2005; Thompson et al., 2012; Zarate et al., 2013), and thus differences between the two types of perception may also extend to the pitch processing mechanisms that underlie them. Thompson et al. (2012) have shown that changes in sound intensity level, which affect the stimulus energy level but not the stimulus waveform, alter the perceived size of ascending melodic intervals and introduce an interval size illusion. In their study, if the higher tone of a 6-semitone interval was louder in sound intensity than the lower tone, then the interval seemed larger than a comparison 7-semitone interval that had no such level variations. This was true for both discrete musical intervals and frequency sweeps. In a related study of subjective interval size, Russo and Thompson (2005) were able to create similar interval illusions by changing the timbre of one of the interval’s two component tones – a
manipulation which changes the distribution of energy in the signal. If the higher tone of a melodic interval was brighter in timbre than the first (i.e. it had more energy at high frequencies than at low frequencies), the interval seemed larger than it was; inversely, if the lower tone had a duller timbre (greater energy at lower frequencies) then the interval seemed smaller than it was. The authors furthermore showed that these timbre manipulations did not distort the pitches of the individual tones of each interval. This effect was present in both musicians and nonmusicians, although musicians were much less susceptible and were only affected by descending intervals (F. A. Russo & Thompson, 2005). Zarate et al. (2013) quantitatively documented the same effect using interval discrimination thresholds and showed that changing the timbre of one of an interval’s two tones led to increased interval-discrimination thresholds. Nonmusicians were especially affected by the changes, and showed a greater bias towards selecting the timbre-altered interval as the larger interval (Zarate et al., 2013). In each of these cases, the timbre and intensity changes affected the perceived size of pitch intervals but did not affect the perceived pitches of the tones that comprised the interval, and both musicians and nonmusicians were affected – though musicians often less so. Taken together, these results suggest that interval perception and basic pitch perception are differentially affected by changes to the physical stimulus properties.

The results mentioned previously also introduce another possibility, which is that different listeners – musicians versus nonmusicians – use different listening strategies, as suggested by Zarate et al., (2013). In both the above studies, nonmusicians were more affected by stimulus manipulations than musicians, although the effects still exist for musicians too. Therefore, musicians and nonmusicians may use different listening
strategies that differentially rely on phase-locking versus energy cues. The well-documented differences between musically trained and musically untrained listeners at both psychophysical (Kishon-Rabin et al., 2001; Micheyl et al., 2006) and physiological levels (Bidelman, Gandour, & Krishnan, 2011; Bidelman & Krishnan, 2010; Lee, Skoe, Kraus, & Ashley, 2009; Wong, Skoe, Russo, Dees, & Kraus, 2007) provide evidence for a potential difference in the listening strategies of these two populations.

Musical pitch perception is thought to be based on temporal pitch mechanisms. However, since each of the previously discussed studies used some form of spectral and energy manipulation, which should theoretically leave phase-locking cues unaffected, the findings suggest that an energy-based strategy may be involved in interval perception, particularly for nonmusicians. While it’s likely that both spectral and temporal cues contribute to pitch-interval perception, few studies have documented additional factors that may influence perception beyond timbre, intensity, and standard interval size. A more thorough understanding of the factors affecting pitch-interval perception would improve our understanding of the mechanisms of relative pitch perception and the contribution of temporal pitch processing. The study presented in this chapter seeks to fill that gap, using a stimulus manipulation that may help improve our understanding of the role of phase-locking and temporal cues in interval perception: frequency region. As mentioned previously, frequency region provides a useful way of examining perception in contexts with degraded phase-locking as a result of the lack of phase-locking at high frequencies. However, few psychophysical studies have directly examined the role of temporal pitch cues in interval perception specifically.
The goal of the current study is thus twofold: to investigate the contribution of phase-locking cues in pitch-interval perception, and to examine whether this integration of energy-phase-locking cues differs between populations with different auditory experiences. To investigate the role of phase-locking cues, subjects are tested in an interval-discrimination task at both low and high frequencies. Due to the nature of musical stimuli and the stimulus roving necessary to ensure the task is not performed by simple frequency discrimination, we are unable to test stimuli at frequencies entirely above 5 kHz, and instead use stimuli that are between approximately 1500 and 5 kHz. If performance degrades at high frequencies, as has been seen elsewhere using non-musical stimuli (Moore, 1973; Sek & Moore, 1995), then a temporal-based mechanism may be involved in interval perception (recall cut-off frequency for phase locking in humans, which begins at approximately 2 kHz). Since performance in other temporal-based listening tasks are known to deteriorate beginning at only 1.5 kHz, we expect that this limited lower-frequency region may still be informative despite not exceeding conservative estimates of the upper limit of phase-locking in humans.

To examine whether musicians and nonmusicians might differ in their perception of such stimuli, we incorporated listeners with three different levels of musical training – untrained, experienced, and formally trained. Like in Chapter 3, we expect that expert musicians may show better performance in general and potential differences in the pattern of performance between the three subject groups might furthermore indicate differences in pitch perception strategies.

Lastly, the experiments discussed in the current chapter are a direct extension of the relative pitch experiments discussed in Chapter 3, and thus the methods are the same, with
the exception of the stimuli at higher frequencies. For clarity, the relevant details are included here.

2. Methods

2.1 Subjects

The data from Chapters 3 and 4 were collected concurrently, and the same subject pool was used. All fourteen adult listeners who participated in the study had normal hearing, none had Absolute Pitch, and none spoke a tonal language. Subjects were classified into three groups based on their degree of musical experience. Five nonmusicians (all males, M = 25.6 years of age, SD = 3.7 years) had less than three years of musical instruction during childhood. Three of the nonmusicians had no musical experience and two had 1 and 3 years respectively (M = .75 years experience, SD = 1.3 years experience). Three amateur musicians (1 male, M = 23.7 years of age, SD = 2.9 years) had between 10 and 12 years of music lessons but no formal music theory (M = 10.67 years experience, SD = 1.2 years experience). All amateur musicians reported playing music recreationally within the past month, and were considered amateur because they had never studied music theory or received ear training. Six expert musicians (4 males, 2 females, average age 26.2 years, SD= 5.6 years) were currently studying or had studied music as part of an undergraduate music program, had at least 12 years of music lessons and 1 years of ear training (M = 16.8 years of musical experience, SD = 6.7 years of experience). All expert musicians were actively engaged in musical composition and four of the six practiced music on a weekly basis. Of the six expert musicians, two were classically trained in both piano and violin, two were classically trained in violin and piano and guitar, one was a jazz violin
and mandolin player, and one was a jazz drummer. Participants were financially compensated for their time and all procedures were approved by the University of California, Irvine’s Institutional Review Board.

2.2 Stimuli

Stimuli and equipment are similar to those used in chapter 3. Tones 1 and 2 comprised the trial interval A and tones 3 and 4 comprised trial interval B. All tones were 500 ms in duration with 20 ms on/off ramps. The two tones of each interval were separated by a 250 ms gap and the two intervals of a trial were separated by a 1000 ms gap. All stimuli were generated in Matlab (Natick, MA) and played over Sennheiser HD380 pro headphones at a sampling rate of 44.1 kHz. To ensure that every listener heard all tones of the experiment at the same sound level, stimuli were inverse filtered with the headphone transfer function to generate an at-ear sound level of 70 dB SPL for all tones.

All intervals were ascending melodic intervals, where the second tone of each interval was higher in frequency than the first. The two intervals of a trial were always unequal in magnitude, and subjects were instructed to determine which was larger. The size of the two intervals of each trial is given by $i$ and $i + \Delta i$, where $i$ is the magnitude of the smaller standard interval (fixed within a run) and $i + \Delta i$ is the magnitude of the larger comparison interval. $\Delta i$ is the difference in magnitude between the two intervals and was determined in an adaptive staircase (see below). The order of presentation of the larger interval was counterbalanced across trials.

To ensure that subjects were not able to perform the task by a simple frequency-comparison of the two highest tones of each interval, the tones 1 and 3 were roved in a
continuous range of ±9 semitones around a central frequency $F$, and were always at least 4 semitones apart.

### 2.3 Adaptive Procedure

The difference in magnitude between the two intervals, $\Delta i$, was controlled by a 2-down-1-up adaptive procedure that converged on 70.7% performance (Levitt, 1971). At the start of each run, $\Delta i$ was initialized to 12 semitones and was decreased after two consecutive correct trials, and increased after 1 incorrect trial (see below for details). At each step, $\Delta i$ was adjusted according to the equation:

$$\Delta i = 10^{\log_{10}(\Delta i) \pm y}$$

where $y = 0.2$ up to and including the 4th reversal, and $y = 0.05$ thereafter. Runs ended after the 50th trial. At the conclusion of each run, the threshold (in semitones) was calculated from the geometric mean of the value of $\Delta i$ at all reversals except the first 3 (for runs with an odd number of reversals) or the first 4 (for runs with an even number of reversals).

### 2.4. Experimental design and Procedure

The experiment followed a $3 \times 3 \times 2$ mixed design. Three different interval-magnitude standards of $i = 1$, 6-, and 7-semitones were tested in two different frequency ranges, yielding a total of six experimental conditions. In the low frequency range, the base tone of each interval was roved around a center frequency $F = 220$ Hz, and all tones in the condition were between 130.8 Hz and 1105.3 Hz. The high frequency region had a center frequency $F = 1760$ Hz and all tones fell between 1046.6 Hz and 8853.5 Hz.
The experiment was administered in two separate 1-hour sessions conducted on consecutive days. Subjects completed 3 conditions per day. With the exception of differences in experimental conditions, all experimental procedures were analogous to those of Chapter 2 and are excluded here.

3. Results

3.1. Effects of musical experience

Figure 4.1 shows the mean pitch interval discrimination thresholds for the three subject groups in both the low and high frequency regions. A mixed-ANOVA shows a significant main effect of group, $F(2,11) = 6.389$, $p = 0.014$, with expert musicians producing the lowest thresholds of the three groups. These findings agree with previous studies demonstrating that musicians perform better than nonmusicians on pitch- and interval-discrimination tasks (Kishon-Rabin et al., 2001; McDermott, Keebler, et al., 2010; Micheyl et al., 2006; Spiegel & Watson, 1984; Zarate et al., 2012, 2013).

Independent t-tests performed on the mean thresholds of each subject reveal that there was a significant difference between the thresholds for the non-musicians and the expert musicians [$t(9) = 3.14$, $p = 0.012$], but only a marginal difference between the amateur musicians and the expert musicians [$t(7) = 2.238$, $p = 0.060$]. There was also no significant difference between the thresholds for the amateur and non-musicians [$t(6) = 0.805$, $p = 0.451$]. Amateur musicians performed similarly to nonmusicians, even in the high frequency regions [$t(6) = 1.463$, $p = 0.194$], a finding inconsistent with previous studies of musical trained listeners that showing differences between musicians and nonmusicians in pitch and interval-perception tasks. This may be caused by the high
variance of both the nonmusicians’ and amateur musicians’ results; when comparing only the thresholds of the 7-semitone condition, in which nonmusicians show less variance, independent t-tests show a marginal difference between the two groups, $t(4.848) = 2.542$, $p = 0.053$.

![Graph of mean thresholds of three subject groups for standard interval distances of 1-, 6-, and 7-semitones in both frequency regions. The left panel shows the low frequency region and the right panel shows the high frequency region. Thresholds are lowest for a 1-semitone standard interval than for 6- or 7-semitone standard interval magnitudes. Expert musicians produced the lowest thresholds of all three groups, and nonmusicians produced thresholds that differed between frequency regions. Error bars represent 1 standard error of the mean. Note that lower thresholds indicate better performance in the discrimination tasks.]

Figure 4.1. To compare with previous studies, a one-way ANOVA was run on the 1-semitone condition and showed a strongly significant main effect of musicianship [$F(2,11) = 8.284$, $p=0.006$], with nonmusicians again producing the highest thresholds and expert musicians the lowest.
3.2. Effects of frequency region

Figure 4.2 shows median thresholds of three subject groups for each frequency region. A 3-way mixed ANOVA shows a significant main effect of frequency region, $F(1,11) = 11.703, p = 0.006$, with listeners producing higher thresholds in the high frequency region.

![Figure 4.2. Boxplot of median thresholds for each group in both low and high frequency regions. Red lines indicate median thresholds value in semitones, and blue box denotes the 25 to 75th percentiles (excluding statistical outliers). Superimposed open black triangles, squares, and circles on the graph indicate individual data from 1, 6, and 7 semitone conditions. Red asterisks indicate statistical outliers. Nonmusicians are highly affected by frequency region manipulations, producing higher pitch-interval discrimination thresholds at high frequencies relative to low frequencies. Amateur and Expert musicians do now show such a pattern of results.

There was also a significant two-way interaction between frequency region and group, $F(2,11) = 8.128, p = 0.007$. Planned comparisons show that only nonmusicians are affected by frequency region, $F(1,4) = 84.873, p = 0.001$, while amateur and expert
musicians were not [F(1,2) = 0.523, p = 0.545 and F(1,5) = 1.442, p = 0.284, respectively]. In other words, nonmusicians were more adversely affected by frequency region and showed a greater difference between the low and high frequency regions than either amateur musicians or expert musicians.

3.3. Effects of standard-interval size

In agreement with the results of Chapter 3 and those of Luo et al (2014), thresholds again varied with standard intervals size and were lower for 1-semitone standard than either 6 or 7-semitone standards, F(2,22) = 28.927, p < 0.001.

4. Discussion

The goal of the current study was to investigate how temporal phase-locking cues contribute to the perception of pitch-intervals, and how the use of these cues may be affected by musical training. To do this, we tested nonmusicians, amateur musicians, and expert musicians in a pitch-interval discrimination task using pitch-interval stimuli presented at both low and high frequency regions.

We found that overall, musicians perform better than nonmusicians, in agreement with previous studies that demonstrate differences between musicians and nonmusicians (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004; Kishon-Rabin et al., 2001; Lee et al., 2009; Micheyl et al., 2006; N. M. Russo, Nicol, Zecker, Hayes, & Kraus, 2005). Contrary to these studies, but in agreement with the results of Chapter 3, results also show that amateur musicians did not produce thresholds that were significantly different from nonmusicians in the task.
Although nonmusicians produced thresholds that were not significantly lower than those of nonmusicians’ overall, they demonstrated a pattern of results with regard to frequency region: while the nonmusicians were affected by frequency region and producing significantly higher pitch-interval discrimination thresholds when discriminating between two intervals in the high (>1.5 kHz) region, neither the amateur musicians nor the expert musicians were affected in this way. Put differently, the pitch-interval discrimination performance of listeners with musical training was not affected by the frequency region of the stimuli, suggesting that they either did not rely heavily on phase-locking cues or that they did not have the same level of phase-locking degradation that the untrained listeners did.

These data do not easily support the suggestion made by previous researchers (F. A. Russo & Thompson, 2005; Thompson et al., 2012; Zarate et al., 2013) that musicians are less likely to use spectral cues; instead it suggests two things: (1) musicians are more likely to use temporal cues (or to have a dual-pitch processing mechanism in relative pitch), or (2) musicians are better able to use the phase-locking cues at low frequencies than nonmusicians. Although further studies are needed to more thoroughly investigate this research question, it’s worth noting that stimuli in the high frequency regions had frequencies as high as 5000 Hz, which is clearly within the region where phase-locking is thought to be heavily degraded in even the best listeners. Therefore, it’s highly likely that at least some of the tones were presented in regions where phase-locking was highly degraded. Either way, these data indicate that there is indeed a different in listening strategy between musically experience and musically naïve listeners, with musically
inexperienced listeners more likely to be affected by changes in the frequency region of the pitch intervals.
CHAPTER 5:
Summary and Conclusions

In conclusion, the studies presented in this dissertation investigate the factors affecting relative pitch perception in a variety of contexts. The experiments presented in Chapter 2 investigate how a listener's ability to judge the direction of frequency sweeps, a percept critical for tonal language perception, is affected by the rate and extent of frequency change. This paradigm is less often studied than other forms of relative pitch perception for continuous tones, such as frequency modulation detection or frequency change detection, and the results suggest that listeners extract the start and end pitches of each sweep and then compare them to determine sweep direction. This type of mechanism, called a sampling mechanism, was originally proposed by Hartmann and Klein (1980) for the detection of frequency modulation, and has since been applied to many other related contexts of relative pitch perception. The data do not suggest a mechanism based on simply detecting changes at the output of auditory filters because this mechanism may not directly provide information about the direction of change (Lyzenga et al., 2004). Furthermore, these two experiments not only evaluated the role of a sampling mechanism in direction identification perception, but are also the first also test the role of temporal processing cues in a sampling mechanism. By comparing performance between low and high frequency regions in which temporal phase-locking cues are available and unavailable, I suggest that sampling mechanisms are at least partially based on temporal pitch processing at low frequencies.
In Chapters 3 and 4, I investigate the factors affecting the perception of pitch distance for discrete changes in pitch, or pitch-intervals. Chapter 3 highlighted a novel effect of standard-interval magnitude, which was present in all experimental groups but formally documented in only one other study (Luo et al., 2014). Because this effect is not often seen elsewhere and performance may seem surprisingly poor – experts’ discrimination thresholds often exceeded 1 semitone – I detail reasons for the source of these results and highlight other interval-discrimination results (McDermott, Keebler, et al., 2010) which may corroborate the results found here. There are also numerous contexts in which trained musicians fail to perform perfectly in musical tasks (Dowling, 1978; Dowling & Fujitani, 1971; F. A. Russo & Thompson, 2005; Siegel & Siegel, 1977b; Zarate et al., 2013), allowing for the possibility that musicians’ may not perform as well as expected in these pitch interval discrimination tasks.

Chapter 4 introduced the factor of frequency region, a manipulation which takes advantage of the frequency limitations of phase-locking and which is often used as a psychophysical method for evaluating temporal processing. Here I show that nonmusicians are unique in their response to changes in the frequency region of stimuli, and produce significantly higher thresholds in higher frequency regions. Although it is possible that the nonmusicians’ vulnerability to high frequency intervals might arise from their inexperience in quantifying and comparing intervals, this does not explain why these deficits would be limited only to frequencies above 1 kHz. Furthermore, the amateur musicians – who have experience playing music but do not have formal music theory training or experience in interval recognition – were similarly affected by standard interval size and produced thresholds on par with the nonmusicians’, but did not show an affect of frequency region.
This may suggest that simply being inexperienced in musical interval tasks was not the only source of the nonmusicians’ performance in the high frequency region.

In sum, the experiments presented in this dissertation highlight some of the factors affecting how we perceive relationships between pitch changes. I studied both continuous frequency changes and discrete frequency changes, and suggest that, at low frequencies where phase-locking is well documented, temporal pitch processing mechanisms may underlie relative pitch perception.
REFERENCES


http://doi.org/10.1177/025576148400300101


1949.


Moore, B. C. J. (1973). Frequency difference limens for short-duration tones the minimum limitations will be determined An alternative type of model assumes investigation nerve provide similar of jitter in individual can be minimized ), could. Journal of the Acoustic Society of America, 54(3), 610–619.


http://doi.org/10.1073/pnas.1015291108


http://doi.org/10.1016/0013-4694(94)00209-4


http://doi.org/10.1016/0013-4694(88)90211-8


nuclei of the chicken: tonotopic organization of n. magnocellularis and n. laminaris.


http://doi.org/10.1002/cne.901640403


http://doi.org/10.1016/0304-3940(91)90645-A


http://doi.org/10.1371/journal.pone.0075410


http://doi.org/10.1038/nature01743.1.