Perception of multiple pitches: Sequential and simultaneous pitch relationships

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Jackson Everett Graves

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Andrew J. Oxenham, Ph.D.

January 2018

Acknowledgements

I am deeply grateful to my advisor, Andrew Oxenham, for his patience, wisdom, and constant encouragement over many years and projects together. I feel very lucky to have had a mentor so knowledgeable, so willing to dedicate time and effort to his students, and so pleasant to work with. I also thank my other committee members: Daniel Kersten, Peggy Nelson, and Yuhong Jiang, for their insightful comments on the dissertation, and invaluable mentorship throughout my graduate career.

A special thanks to Christophe Micheyl, whose encouragement and advice were instrumental for me both early in Minneapolis and later in Lyon, and who directly contributed to statistical work included in this dissertation.

I could never have finished this project without the support of my family, especially my parents, and the companionship of many dear friends who contributed in important ways. Special mention is due to:

- Anahita Mehta, who pushed me relentlessly to do my best, who commiserated with me when things went poorly and rejoiced with me when I succeeded, and who was there with me at all the most crucial moments of the process, even at unholy hours of the night.
- Kelly Whiteford, who influenced this project, among other ways, through countless short and spontaneous conversations across the divider between our adjacent desks, and whose opinion I value as highly as her friendship.
- Elizabeth Fast, a fellow traveler through graduate school from the beginning, who opened her home to me and in whose guest bedroom much of this document was written.
- Marion David, who provided insightful advice over the course of many years, et qui m'a aidé énormément dans mes efforts à apprendre le français.

I'm lucky to have had the chance to work alongside all of the other members of the Auditory Cognition and Perception laboratory at the University of Minnesota as well, a group of excellent scientists and wonderful colleagues, including Magdalena Wojtczak, Heather Kreft, Emily Allen, Jordan Beim, Coral Dirks, and many others. I thank undergraduate research assistants Alex Oster and John Koch for help in collecting behavioral data.

I also owe a huge debt of gratitude to my mentors in Lyon, Barbara Tillmann and Anne Caclin, who taught me new experimental and analytical techniques and new ways to think about data, and who gave me a warm welcome during my visit, as well as many other colleagues at the Lyon Neuroscience Research Center.

This work was funded by NIH grant R01 DC005216, as well as a grant from the Erasmus Mundus Student Exchange Network in Auditory Cognitive Neuroscience, and the LabEx CeLyA and Cortex (ANR-11-IDEX-0007).

Abstract

The perception of pitch, a dimension of sound that is important for music perception, speech perception, and sound source segregation, is influenced by its context, both sequential and simultaneous. In music, pitch sequences form melodic contours, and simultaneous pitches form chords and harmony. A series of experiments investigated the perception of melodic contour in pitch as well as two other auditory dimensions, brightness and loudness. The results showed that subjective ratings of continuation for brightness and loudness sequences conformed to the same general contour-based expectations as pitch sequences, suggesting that melodic expectations are not unique to the dimension of pitch. Listeners with congenital amusia, however, exhibited less impairment on a short-term memory task for loudness contours than for pitch contours, suggesting a pitch-specific deficit. In a pair of experiments, priming of a familiar tonal context improved accuracy on a pitch interval discrimination task. However, the overall benefit to performance from tonal context was small, suggesting that previously reported effects of response time may mainly reflect expectancy as opposed to perceptual accuracy. In the last series of experiments, listeners accurately identified pitches in mixtures of three concurrent complex tones, despite poor peripheral resolvability. These stimuli help to dissociate two normally confounded variables in complex pitch, harmonic number and peripheral resolvability. The results were compared with outputs from two kinds of auditory models, one based on the rate-place code for pitch and the other based on the temporal code. Overall, these findings suggest that pitch perception involves bottom-up integration of both spectral and temporal information, as well as top-down effects of learning and context.

Table of Contents

List of Figures and Tables

List of Abbreviations and Acronyms

ACF: Autocorrelation function

AN: Auditory nerve

CF: Center frequency

DL: Difference limen

F0: Fundamental frequency

LTM: Long-term memory

NCCF: Normalized cross-correlation function

RT: Response time

SACF: Summary autocorrelation function

ST: Semitone

STM: Short-term memory

TEN: Threshold-equalizing noise

TFS: Temporal fine structure

Chapter 1: Introduction to pitch perception

I. General overview

Pitch is a perceptual dimension along which periodically repeating sounds can be ranked from low to high. It is strongly correlated with the overall repetition rate of a sound. Various definitions of pitch have been proposed, but one generally accepted definition states that pitch is "that attribute of sensation whose variation is associated with musical melodies" (Plack, Oxenham, Fay, & Popper, 2005). As this definition suggests, pitch is an important dimension for music perception. It defines both melody and harmony, two crucial elements of Western music. However, it is worth noting that pitch is not the only auditory dimension that influences perception of melodies: for example, the dimensions of brightness and loudness can also be used to produce melodic contours which allow for melody recognition and memory (McDermott, Lehr, & Oxenham, 2008). Although it is often defined in connection with music, pitch is also a useful cue for many non-musical auditory tasks. In speech perception, pitch conveys emotion, emphasis, and other prosodic information, as well as semantic meaning in tone languages. In various listening situations, pitch can be a useful cue for sound source segregation: for example, when listening to simultaneous speech sounds, a pitch difference helps to perceptually segregate the two voices (Bregman, 1990, p. 559). In any of these listening contexts, the process of perceiving pitch may subjectively seem effortless. However, like many other seemingly simple perceptual tasks, this computational process is in reality very complex, and it has proven challenging to understand and model the auditory system's methods of accomplishing it.

Although researchers in laboratory settings may present isolated tones that evoke isolated pitch percepts, most pitched sounds in everyday life occur in context, and are often defined mainly in relation to other pitches that have occurred recently, or that occur simultaneously. The importance of relative pitch over absolute pitch is observable in humans from an early age (Plantinga & Trainor, 2005), suggesting it may be either innate or learned early, but is more elusive in other species (Bregman, Patel, & Gentner, 2016; D'Amato, 1988; Yin, Fritz, & Shamma, 2010), suggesting it may be unique to the perceptual needs of humans. In any case, our perception of pitch patterns unfolding over time is sophisticated, allowing for both short-term and long-term storage of specific pitch patterns (Dowling & Fujitani, 1971; Dowling, 1978). Novel pitch sequences even induce specific expectations for their continuation over time (Cuddy & Lunney, 1995).

In addition to pitches presented in sequence, human listeners are generally able to perceive multiple simultaneous pitches without great difficulty. In music, the presence of three or more simultaneous pitches is the rule, not the exception. However, existing computational algorithms for the estimation of a single pitch (de Cheveigné & Kawahara, 2002; Noll, 1967) are far more effective than algorithms that attempt to estimate multiple simultaneous pitches (e.g. de Cheveigné & Kawahara, 1999; Klapuri, 2008; Yeh, Roebel, & Rodet, 2010). A great deal of psychoacoustic research has been conducted on the perception of single pitches, but comparatively little exists on the perception of multiple simultaneous pitches, despite the prevalence of listening situations involving multiple simultaneous pitches in everyday life.

The present dissertation presents an exploration into human perception of pitch relationships, both sequential and simultaneous. To provide some background for these studies, the following section reviews some basic concepts and findings associated with the perception of pure tones and harmonic complex tones.

II. Pitch of pure tones

A pure tone is a theoretical waveform with pressure repeating at a given frequency over time in a perfectly sinusoidal pattern. Pure tones generally do not occur in the natural environment, and even artificially generated stimuli referred to as “pure tones” are not technically pure, as they are not infinite in duration, containing onsets and offsets which theoretically introduce other frequencies. Despite their lack of direct ecological relevance, pure tones are highly useful as a simple case to test the responses of the auditory system to single frequencies.

Even with stimuli as simple as pure tones, it is readily apparent that the perceptual dimension of pitch does not correlate perfectly with any single physical dimension of sound. A pure tone is fully described by its frequency, phase, and amplitude, and it is frequency that correlates most strongly with perceived pitch, but there are complications even to the relationship between the frequency and pitch of a pure tone. First of all, the relationship is not linear – it approximates a logarithmic scale, such that a given frequency ratio produces a fixed magnitude of perceived pitch change. This is the way pitch distances are labeled in music: for example, the pitch distance of an octave always corresponds to a doubling in pure-tone frequency. Some early studies have tried to psychophysically measure the relationship between pitch and pure-tone frequency (Siegel, 1965; Stevens, Volkmann, & Newman, 1937), finding results that approximate, but do not exactly match, this logarithmic relationship defined in music theory. These efforts resulted in the Mel scale (Stevens et al., 1937) that has been used in some automatic speech recognition applications (Farooq & Datta, 2001), but has failed to find acceptance in the pitch literature (Attneave & Olson, 1971). Even with perfect knowledge of the relationship between frequency and pitch of pure tones, knowing a pure tone's frequency alone does not tell us everything about its pitch. Early studies also found that the amplitude of a pure tone has a small, but measurable, effect on its pitch, such that high frequencies (above about 2 kHz) seem to have higher pitch at higher intensities, while low frequencies (below about 2 kHz) seem to have lower pitch at higher intensities (S. S. Stevens, 1935; Verschuure & van Meeteren, 1975).

According to Fourier's theorem, any sound may be described either in the time domain, as pressure over time, or in the frequency domain, as the magnitudes and phases for each frequency component within a sound. One unique quality of a pure tone is the exact equivalence between its overall temporal repetition rate and the location of the single peak in its magnitude spectrum – these are both equal to its frequency. From the point of view of an auditory system trying to extract the pitch of a periodic sound, this means that for a pure tone, any pitch extracted using the spectrum will be identical to the pitch extracted using the time domain.

The physiology of the human auditory system allows, in theory, for both of these strategies. The strategy that uses the frequency spectrum has been called the rate-place representation: different places along the length of the basilar membrane in the cochlea have different resonant frequencies, and so a pure tone at a given frequency will produce a peak motion, and hence peak neural firing rate, at one specific place in the cochlea. The strategy that uses the time domain can be called the temporal representation: all the neurons that respond (at any firing rate) will "phase lock", tending to fire in synchrony at a particular phase in the cycle, producing spikes in the population firing rate at regular time intervals equal to the period of the waveform (the inverse of its frequency), at least at low frequencies.

Researchers have used pure tones to study the relative efficiency or usefulness of these two possible codes for pitch in different spectral regions. When difference limens (DLs) are measured for the frequency of pure tones, they are found to dramatically increase as the frequency being discriminated exceeds about 5 kHz (Moore, 1973). This result has been explained in terms of the upper limit of phase locking, such that the temporal code is used for low-frequency pure tones, and the rate-place code is used for high-frequency pure tones. Phase locking, the basis of the temporal code, has a theoretical upper limit in frequency: at the point where the temporal noise or error in phase locking exceeds the period of the waveform, an accurate temporal representation is no longer be possible. This lack of access to the temporal code at the high frequencies leaves the system dependent on the rate-place code, which could be the explanation for elevated DLs above 5 kHz.

Another notable change in pitch perception for pure tones above 5 kHz is that this seems to be, in some sense, the upper limit of "musical pitch." Below this limit, when listeners are asked to subjectively match the size of a pure tone interval to a reference interval in a different spectral region, they match the interval size according to the musical frequency scale, an exact logarithmic relation of frequency to pitch (Attneave & Olson, 1971), as opposed to the Mel scale measured psychophysically (Stevens et al., 1937). Above about 5 kHz, pure tone interval size matching behavior becomes erratic, suggesting that listeners lose the ability to accurately perceive or represent melodies.

Although it cannot be directly measured in humans, this limit has been assumed to be the point at which phase locking breaks down, preventing the use of the temporal code for pitch. Another way in which 5 kHz functions as a limit for musical pitch is as roughly the upper limit of pitches produced by musical instruments: the highest note on a grand piano is near this frequency, as well as the highest note for the highest-pitched instruments in an orchestra, such as the piccolo.

III. Complex pitch

Complex tones, or sounds that contain more than one frequency, are a more ecologically valid stimulus than pure tones, as most pitches in the natural environment, such as animal voices and musical instruments, are produced by complex tones. Specifically, these sounds in the environment take the form of harmonic complex tones, meaning they are composed of frequency components at integer multiples of the fundamental frequency (F0). The pitch of a harmonic complex tone corresponds to its F0, even when the frequency component at F0 itself is absent. In the 19th century, the pitch of a complex tone was a point of controversy, with Ohm (1843) arguing that the pitch was determined by the lowest harmonic component, and Seebeck (1841) arguing instead that the pitch was determined by the F0. Ultimately the most conclusive evidence that complex pitch corresponds to F0 came from the finding that even when the component at F0 is removed and masked by low-pass noise, the resulting sound has a pitch corresponding to the missing F0 (Licklider, 1956). The control measure of masking the spectral region around F0 with noise was necessary to prevent the possibility that F0, after being removed, could be reintroduced in the ear by nonlinear distortions in the response of the cochlea.

Beyond the F0, some of the lower components in a harmonic complex tone can also be removed and masked with noise without weakening the pitch percept resulting from the remaining, higher components. However, when increasingly high harmonic components are removed, above about the 10th multiple of F0, the salience of the pitch percept significantly weakens (Houtsma & Smurzynski, 1990). For high-harmonic-only complexes like that, listeners are less accurate when labeling musical intervals, and show a marked elevation in DLs for F0, indicating poorer discriminability. For both of these tasks, a steep decline in performance is observed between complexes including components as low as the 7th (where performance is good) and complexes including no components below the 13th (where performance is poor). Outside of this transition region between the 7th and 13th harmonic, however, lowest included harmonic number seems to make little difference to performance in pitch perception tasks. It should be noted that in some cases, pitch based only on the temporal envelope (evoked by amplitude-modulated broadband noise) can also be strong enough to support labeling of musical intervals (Burns & Viemeister, 1976).

A possible explanation for this sharp weakening of the pitch percept when only harmonics above about the 10th are included is that harmonics in this region are unresolved: they are spaced so closely together that they do not produce separate peaks of excitation that can be evaluated with the rate-place code. High-numbered harmonics are more likely to be unresolved than low-numbered harmonics, due to the fact that cochlear filtering is approximately logarithmic, meaning that bandwidths increase with increasing center frequency (Glasberg & Moore, 1990), whereas harmonic spacing remains constant in absolute frequency.

This explanation seems to be supported by findings of phase effects for complexes with only high harmonics. Adding components together in the special Schroeder-phase relations (Schroeder, 1970), which minimizes the peak ratio of the temporal envelope of the complex, results in higher DLs than adding the components together in the maximally peaky sine-phase relation (Houtsma & Smurzynski, 1990). But this is only true for high-harmonic complexes, suggesting that the temporal waveforms of the high harmonics are being summed together as more than two components pass through each auditory filter. Only for unresolved harmonics would the temporal pitch mechanism respond to summed waveforms of multiple components, and thus become sensitive to additive phase relations. Taken together, the elevated F0DLs and the emergence of phase effects for 10th-harmonic-and-up complexes suggest that peripheral resolvability is the reason for the weakened pitch perception for these complexes.

Further evidence pointing towards separate mechanisms for resolved and unresolved pitch comes from a study that used stimuli designed to produce different temporal representations from their rate-place representations (Shackleton & Carlyon, 1994). Listeners had to match the pitch of a reference complex to the pitch of a complex whose components were either added in sine phase or in alternating sine-cosine phase (known as “alt-phase”). Despite consisting of the exact same frequencies at the exact same magnitudes, and thus producing an identical rate-place representation, the temporal envelope for the alt-phase complex repeated at twice the rate of the equivalent sine-phase complex. The perceived pitch of these stimuli is doubled, consistent with the doubling of the temporal envelope repetition rate, but only for high-numbered harmonic complexes. Presumably the high-harmonic complexes were unresolved, leaving only the temporal mechanism to extract the pitch of these complexes. The low-numbered harmonics are resolved, allowing for rate-place based extraction of the original pitch, not based on the repetition rate of the temporal envelope.

Recent research suggests that aging, even in the absence of age-related hearing loss measurable with pure-tone audiometry, can produce a decline in pitch perception (Russo, Ives, Goy, Pichora-Fuller, & Patterson, 2012). In a melodic pitch perception task, younger listeners were better able to detect a pitch shift of one or two semitones in a four-note melody when the melody was composed of complexes including resolved harmonics (from the 4th or the 8th upwards) than only unresolved harmonics (nothing below the 12th). For most older listeners, however, performance did not improve in the 8th-and-up condition relative to the unresolved (12th-and-up) condition, and only mildly improved in the 4th-and-up condition. These results provide new insight beyond previous evidence (Patterson, Nimmo-Smith, Weber, & Milroy, 1982) that frequency selectivity deteriorates as auditory filters broaden with age-related hearing loss. Since all the older listeners in Russo et al. (2012) had audiometric thresholds below 20 dB HL from 250-8000 Hz, this finding suggests that auditory filters may broaden with age even in the absence of clinically significant hearing loss. However, although both groups had low thresholds, the older listeners did have higher thresholds on average than the younger listeners, suggesting that hearing loss may still be ultimately responsible for the decline in frequency selectivity. Regardless, one could interpret the results of Russo et al. (2012) as showing that older listeners had to rely on temporal pitch, rather than rate-place pitch, for even the lower spectral regions, as their broadened auditory filters failed to separately resolve the harmonics even in the lower spectral regions. These findings are consistent with the view that poor resolvability is the cause of worsened pitch perception for high-harmonic missing-F0 complexes.

This view is further supported by findings that in situations where auditory filters are broadened, the minimum harmonic number that must be present to produce low F0DLs shifts downward (Bernstein & Oxenham, 2006a, 2006b). Specifically, when harmonic complexes are presented in a fixed spectral region, such that increasing F0 decreases the lowest present harmonic number, the F0 value marking the transition point from low to high F0DLs shifts upward in conditions that decrease frequency selectivity. In other words, with broader auditory filters, listeners perform poorly at discriminating F0 even for higher F0s where performance would otherwise be good. This correlation between frequency selectivity and F0DL transition point was found in normally hearing listeners by varying stimulus level, which affects auditory filter bandwidth (Bernstein & Oxenham, 2006a). A similar relationship was observed in listeners with sensorineural hearing loss, estimating filter bandwidth using notched noise (Bernstein & Oxenham, 2006b). These findings strongly suggest that resolvability is the driving factor behind the transition from good to poor F0 discrimination.

There is a large weight of evidence supporting the rate-place explanation for the limits of complex pitch perception, but some recent studies have muddied the picture, suggesting high harmonic numbers may produce weak pitch perception regardless of resolvability (Bernstein & Oxenham, 2003, 2008). A more nuanced possibility is that the rate-place and temporal mechanisms may work together to extract complex pitch. One major motivation for a model that combines these two sources of information, rather than just temporal information, comes from behavioral studies using transposed tones, stimuli that contain contradictory temporal and rate-place information (Oxenham, Bernstein, & Penagos, 2004). A transposed tone consists of a high-frequency sinusoidal carrier modulated by a low-frequency half-wave rectified sinusoid. The result is a stimulus that will excite a high-frequency location along the cochlear partition, but which should elicit a functionally similar temporal response to a low-frequency pure tone in the auditory nerve. A model of pitch perception that relied only on the temporal code would predict that the resulting pitch from such a tone would be identical to that of the equivalent low-frequency pure tone. In direct contradiction to such models, listeners' F0DLs for transposed tones are dramatically elevated relative to the equivalent pure tones. When harmonic complex tones were constructed using transposed tones instead of pure tones, listeners were completely unable to perceive complex pitch from these tones, though a purely temporal model of pitch perception would predict an equally strong complex pitch sensation.

In another challenge to purely temporal models, a recent finding has also cast doubt on the widespread assumption that the 5 kHz limit to musical pitch perception is a low-level limitation, arising from the peripheral constraint of the upper limit of phase locking. Instead it may be a learned bias that can be overcome with the right stimuli (Oxenham, Micheyl, Keebler, Loper, & Santurette, 2011). Listeners can hear and discriminate melodies made of complex tones with pitches corresponding to frequencies below the 5 kHz limit, but composed only of frequency components well above that limit. This finding is inconsistent with the theory that phase locking is both necessary for musical pitch and impossible above 5 kHz. One possible conclusion is that it is possible to achieve accurate musical pitch perception without phase locking – perhaps relying in this case on the rate-place code rather than the temporal code.

A further investigation of unresolved pitch using transposed tones found evidence that for unresolved harmonics, listeners may be more sensitive to temporal fine structure (TFS) of a tone than to overall temporal envelope (Santurette & Dau, 2011). The transposed tones used in this study were specifically designed to have two alternating time intervals between TFS peaks that both differed from the regular time interval between envelope peaks. This was accomplished by using frequencies for the carrier and modulator of the transposed tone that were inharmonic with each other. If pitch were dominated by the temporal envelope, these sounds would have a unitary pitch corresponding to the time between envelope peaks (the frequency of the slow, modulating tone of the transposed tone). Instead, listeners matched the pitches of these transposed tones generally in bimodal distributions, around the two pitches corresponding to the two different possible time intervals between local TFS peaks. These results can be explained by a purely temporal pitch model, as long you posit that phase locking is sensitive to TFS over temporal envelope and exists in higher spectral regions than previously assumed.

As an alternative explanation, a further study suggested that for these transposed-tone stimuli, the rate-place code for "unresolved" pitch might be more useful than previously thought – for stimuli at the margin of resolvability, they found listeners were unable to hear out individual harmonics, yet also found no effect of additive phase, the hallmark of unresolved components (Santurette, Dau, & Oxenham, 2012). On balance, the evidence from this and other studies discussed in this section seems to suggest that both rate-place and temporal coding information are necessary/useful for accurate pitch perception, and it seems likely that the method actually employed by the auditory system for pitch perception represents some sort of combined spatiotemporal model.

IV. Overview of chapters

Relative pitch perception for pitch sequences (melodies) has been extensively studied, but a recent study (McDermott et al., 2008) raised many new questions about relative pitch by showing that contour perception is not specific to the auditory dimension of pitch. Chapter 2 addresses the question of whether short-term expectations for continuation in pitch sequences generalize to expectations formed by sequences of sounds varying in loudness or brightness, an aspect of timbre. If we rely on separate mechanisms to form expectations about continuation of sequences in pitch, brightness, or loudness, the kinds of expectations formed may be different in these three dimensions. If we use a common mechanism to form expectations about continuation in all three dimensions, however, the kinds of expectations formed will be the same.

Chapter 3 explores a different question raised by the discovery of contours in brightness and loudness: how do special populations with pitch processing deficits respond to contours in non-pitch dimensions? A group of listeners with congenital amusia (a condition involving deficits in music perception) is compared to a group of listeners without amusia, in terms of their ability to store pitch, brightness, and loudness contours in short-term and long-term memory. If their deficit is specific to pitch, amusics' abilities in the dimensions of brightness and loudness should be less impaired.

Even if contours can be perceived in dimensions other than pitch, one feature that sets it apart as an auditory dimension is the common practice of organizing the pitch scale into hierarchical patterns (scales or keys). Chapter 4 explores the influence of these tonal hierarchies on perception of melodic intervals, in the form of context tones that establish varying degrees of tonality.

Chapter 5 presents three experiments that expand the set of stimuli used in psychoacoustic studies of complex pitch, by evaluating listeners' pitch perception in a mixture of three concurrent complex tones. These experiments are designed to tease apart the effects of spectral resolvability and harmonic number. In Chapter 6, spectral and temporal models are applied to the stimuli from Chapter 5, and the resulting predictions are compared to the behavioral data. Finally, Chapter 7 provides a summary and synthesis of the findings, along with suggestions for future research.

Chapter 2: Sequential pitch: melodic contour in pitch and other dimensions

Sections I-IV are reprinted from:
Graves, J. E., Micheyl, C., & Oxenham, A. J. (2014). Expectations for melodic contours transcend pitch. *Journal of Experimental Psychology: Human Perception and Performance, 40*(6), 2338–47.

I. Introduction to melodic expectation

What makes a good melody? Although the question may be most pressing for composers or songwriters wishing to write the next major hit, it has also been considered from several other perspectives. Cognitive psychologists have noted that when listeners are presented with a sequence of notes, they rapidly form expectations about how the melodic sequence will continue, based either on prior exposure to that melody, or on more general acquired or innate principles (Carlsen, 1981; Cuddy & Lunney, 1995; Huron, 2006). Music theorists have also studied the quality of "good continuation" in melodies, and have developed guidelines for writing perceptually independent melodic lines, referred to as the rules of voice leading (Schenker, 1935).

Studies of melodic expectation have identified two basic categories of expectations, one involving perceived musical key or tonality, and one involving contour – the pattern of directions (up or down) of the intervals in a melody (Narmour, 1990). Although it is necessary to take the influence of tonality into account to provide a complete description of melodic expectation, many of the well-established principles relate only to melodic contour. Novel melodies are more easily distinguished from melodies with different contours than from melodies with similar contours (Dowling, 1978). Indeed, the preservation of melodic contour alone is enough to allow for the memorization of unfamiliar melodies and the recognition of familiar melodies (Dowling & Fujitani, 1971).

Preference and expectation for melodies are distinct concepts (expected melodies may not be preferred), but are closely related, as expected continuations are more likely to be preferred than unexpected continuations. Melodic preferences, particularly those related to tonality, are likely to be culturally specific and so may depend on exposure to certain forms of music and melodies (Kessler, Hansen, & Shepard, 1984; Thompson, Balkwill, & Vernescu, 2000). Other preferences and expectations, particularly those related to melodic contour, may reflect more general perceptual principles related to the formation of auditory streams, and may not be specific to melodies or even music (A. Bregman, 1990; Huron, 2001; Schellenberg, Adachi, Purdy, & McKinnon, 2002). One way to test whether melodic contour expectations are domain specific, or whether they reflect more general perceptual principles, is to generate contours in dimensions other than pitch. Although the concept of melodic contour has traditionally been applied only to melodies consisting of a sequence of tones that vary in pitch, contours can be perceived, remembered, and even used for recognition of familiar melodies in dimensions other than pitch (McDermott et al., 2008).

Pitch is a perceptual auditory dimension primarily related to a sound's overall periodicity or fundamental frequency (F0). The auditory dimension of brightness is an aspect of timbre related to the center of mass of a sound's spectral envelope (sounds with more energy in the high-frequency range of the spectrum are perceived as being brighter). Loudness is primarily related to a sound's intensity. Among these dimensions of sound, pitch is unique in that it can be classified according to both pitch height (a linear scale) and pitch chroma (a circular scale that repeats with every doubling of F0). Furthermore, perceived relationships between pitches form tonal hierarchies: Western listeners, especially those with musical training, judge notes belonging to an established musical scale as better "completions" following that scale (Krumhansl & Shepard, 1979). In the dimensions of brightness and loudness, there are no analogies to pitch chroma or tonal hierarchy, only to pitch height. To the extent that melodic expectations are influenced by tonality, they should not be replicable in other auditory dimensions. However, the aspects of expectation influenced by a melody's contour, which relates only to the linear scale of pitch height, may generalize to domains other than pitch.

In this study we asked whether the same expectations that have been discovered for melodic contours in pitch also apply to contours in brightness and loudness. In two experiments, we presented our participants with 3-tone "melodies" that varied in pitch, brightness, or loudness, and we asked them to judge how well the final note of the melody completed the sequence. Against these results, we tested three well-established rules of melodic continuation, derived from music theory and from cognitive studies based on pitch variations. If expectations for melodic contour extend beyond the pitch dimension, then we would expect listeners' judgments to conform to the predictions of these rules, not only for pitch sequences, but also for sequences based on brightness and loudness. On the other hand, if such expectations are specific to pitch, as expected if melodic contour expectations were learned just from exposure to music, then the rules should only successfully predict the results from pitch-based melodies.

II. Experiment 2.1: Varying melodic context

Method

Stimuli. Harmonic complex tones were shaped with spectral envelopes determined by applying a Gaussian weighting function to the amplitudes of the individual harmonics. The standard deviation of the Gaussian was set to 25% of its center frequency. All the tones were gated on and off with 20-ms raised-cosine ramps. The tones were generated within MATLAB (The Mathworks, Natick, MA) and were played out from a 24-bit L22 soundcard (LynxStudio, Costa Mesa, CA) to both ears through HD580 headphones (Sennheiser USA, Old Lyme, CT), at a sampling rate of 48 kHz. Pitch variations were achieved by varying the F0 of the tones; brightness variations were achieved by varying the center frequency of the Gaussian weighting function; and loudness variations were achieved by varying the overall sound level of the tones. Fig. 2.1 demonstrates the difference between changes in pitch, brightness, and loudness.

The first step in designing the stimuli was to create broadly equivalent "scales" in the three dimensions of pitch, brightness, and loudness. This was achieved by using scale step sizes of 1 semitone (~6%) for F0, 2 semitones for the center frequency of the Gaussian weighting function, and 2 dB for the overall sound pressure level. The step sizes were selected to be approximately equally salient, based on previously reported interval-discrimination thresholds for pitch, timbre, and loudness (McDermott, Keebler, Micheyl, & Oxenham, 2010). It is important to note here that by "scale" we do not mean a musical key or any other kind of tonal hierarchy. Those elements of pitch melodies cannot be meaningfully translated into brightness or loudness melodies, since there is no analog to pitch chroma in those dimensions. “Scale” here simply means a set of ranked, evenly spaced steps from which values for pitch, brightness, and loudness are chosen.

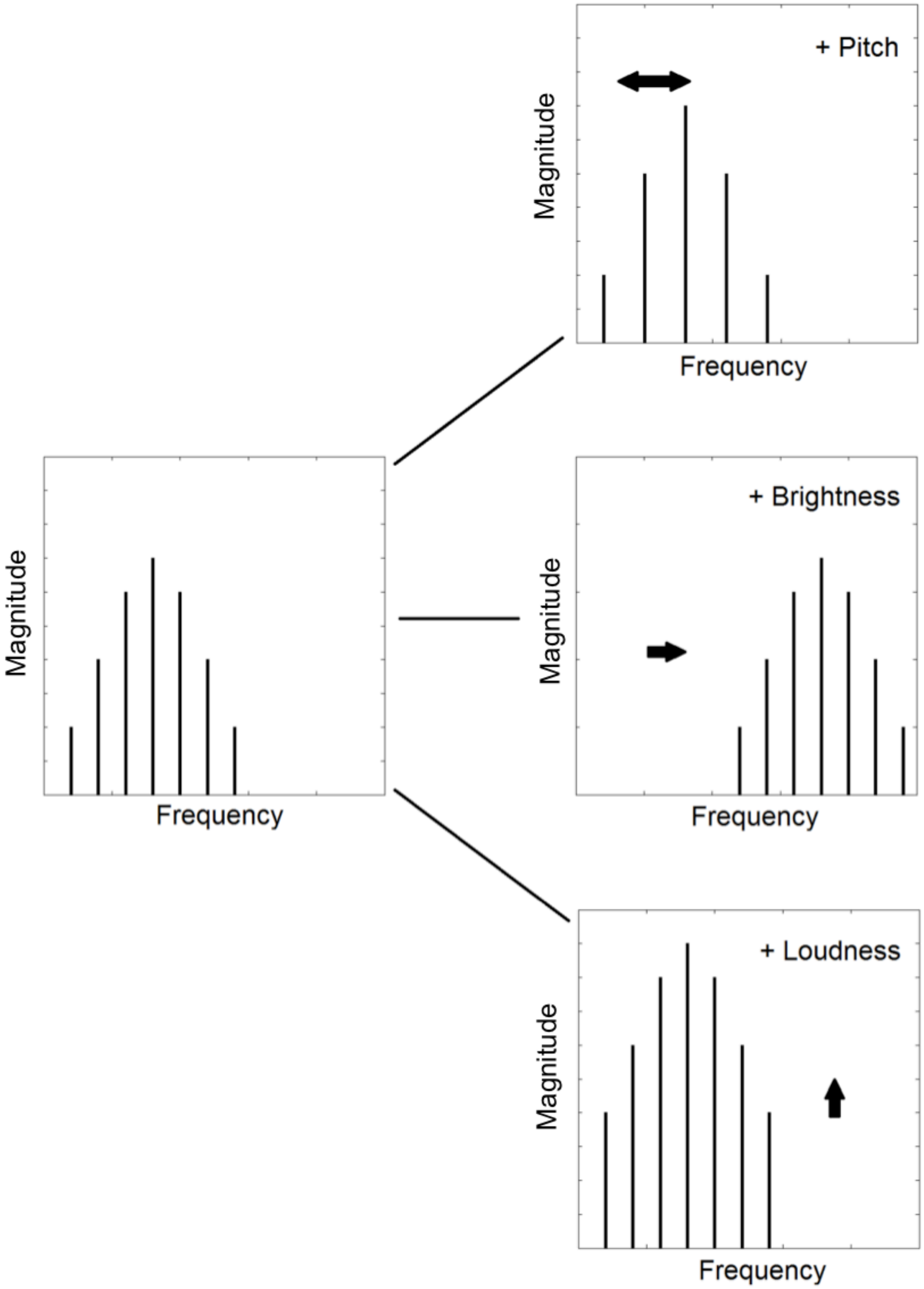


Figure 2.1. Simplified representations of a complex tone (left), increasing in pitch (top right), brightness (middle right), or loudness (bottom right).

The scale for each dimension spanned 27 steps (Fig. 2.2A). In pitch, the F0s ranged from G3 (196 Hz) to A5 (880 Hz) in 1-semitone steps (an equal-temperament tuning including the A440 pitch standard). In brightness, the center frequency of the Gaussian function ranged from 196 Hz to 3951 Hz, in 2-semitone steps. In loudness, the overall level ranged from 30 to 82 dB SPL, in 2-dB steps. The range of these scales was determined by various constraints. First, the minimum and maximum loudness values were chosen to be easily audible and not uncomfortable, respectively. This level range, combined with the step-size of 2 dB, allowed for 27 scale steps. The same number of steps was then used for all three dimensions. The F0 range was selected to span a range that was within that normally used in Western music for melodies. The range of center frequencies for the Gaussian function was selected to begin at the lowest F0, with the highest frequency selected to be 27 steps away, based on a spectral step size of 2 semitones.

Changes in one auditory dimension can interfere with the perception of others (e.g., Borchert, Micheyl, & Oxenham, 2011; Melara & Marks, 1990), so when the stimuli varied along a single dimension, the other two dimensions were held constant. The constant values for the three dimensions were 196 Hz for F0, 800 Hz for spectral center frequency, and 60 dB SPL for sound level. The constant values for spectral center and sound level were selected for their intermediate position in the overall range of values used, while the constant value for F0 was selected to prevent cases where F0 exceeded the spectral center frequency.

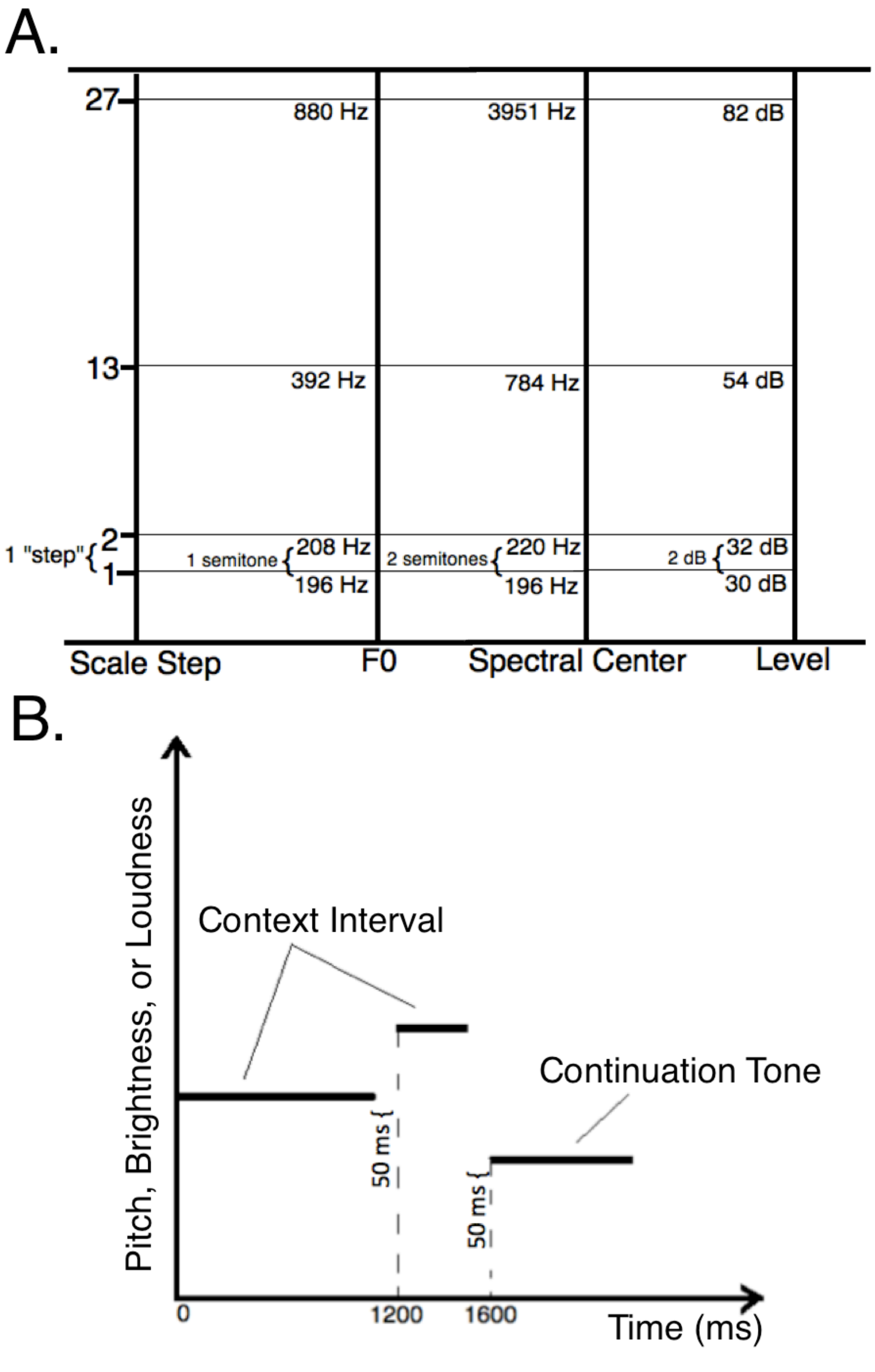


Figure 2.2. (A) Visual representation of the scales used for F0 (for pitch melodies), spectral center (for brightness melodies), and level (for loudness melodies). Each scale contains 27 steps; the values of the 1st, 2nd, 13th, and 27th steps are given as examples. (B) Schematic diagram of an example melody, where horizontal lines represent individual tones in the melody.

Once the scales were established, we adapted a paradigm that was used in an earlier study for generating pitch melodies (Cuddy & Lunney, 1995) to create melodies in the pitch, brightness, or loudness dimension. Melodies consisted of three notes each. The first two notes comprised the context interval. The third note is referred to as the "continuation tone" (Fig. 2.2B). The same eight context intervals originally used by Cuddy and Lunney (1995) were used. In Western music, these intervals in pitch are referred to as the ascending and descending forms of the major second, minor third, major sixth, and minor seventh. These intervals correspond to the following number of steps respectively: ±2, ±3, ±9, and ±10 steps. For each context interval, every continuation tone from 12 steps below to 12 steps above the second tone (25 intervals total) was tested for a total of 200 trials (8 context intervals by 25 continuation intervals). In every melody, the value of the second note was selected from a set of three equally probable values, corresponding to the three centermost values in the pitch, loudness, or brightness range. In pitch, for example, the second note of every melody was randomly sampled from the set of G4 (392.0 Hz), G#4/Ab4 (415.3 Hz), and A4 (440 Hz). The values of the first and third notes were then determined based on the value of the second note and the necessary interval sizes and directions for each trial. To allow for continuation tones 12 steps above or below the second note, 27 different values were defined, and the second note was either the 13th, 14th, or 15th of these 27 values.

The three notes were presented with the temporal relationships shown in Fig. 2.2B. The duration of each note (including onset and offset ramps) was 1150, 350, and 750 ms, respectively. Including the 50-ms silence after each note, the stimulus onset asynchronies were 1200, 400, and 800 ms, which was designed to create a sense of 4/4 meter, with the first and last notes falling on the first beat of the measure (Cuddy & Lunney, 1995). This temporal pattern accents the final note of the melody, which has been shown to heighten performance on perceptual tasks such as pitch change detection for the accented note (Monahan, Kendall, & Carterette, 1987).

Procedure. Eighteen listeners, 5 male and 13 female, were recruited from the Twin Cities campus of the University of Minnesota. Listeners ranged in age from 18 to 31 (*M* = 20.8, *SD* = 3.0). The average amount of musical training was 6.5 years (*SD* = 4.8; range 0 to 13 years). The five participants who reported the lowest amount of musical training (either 0 or 1 years) and the four participants who reported the highest amount of musical training (either 12 or 13 years) were taken as an approximation of the lower and upper quartiles, respectively, of participants ranked by musical experience. All listeners had normal audiometric hearing thresholds (defined as not exceeding 20 dB HL for octave frequencies between 250 and 8000 Hz).

Listeners gave subjective continuation ratings for 200 three-tone sequences each in pitch, brightness and loudness (600 total). After each sequence, the listener was asked to rate how well the third tone met expectations on a Likert scale (Likert, 1932) from -3 ("Very Poorly") to 3 ("Very Well"). Listeners were encouraged to use the full range of possible integer ratings from -3 to 3.

Experiment 2.1 deviated from the paradigm established by Cuddy and Lunney (1995) in two ways. Firstly, the previous study presented the 200 possible melodies in blocks based on context interval size, such that all melodies beginning with the 9-steps-ascending context interval were heard in immediate succession. To avoid possible long-term context effects associated with presenting the same stimulus repeatedly, we randomized the presentation of the 8 different context intervals from trial to trial, just as the presentation of the 25 different continuation tones was randomized from trial to trial. Secondly, Cuddy and Lunney (1995) set the second note of their melodies as equal to C4 or F#4, alternating every other trial. With our selected step size in loudness (2 dB), the range required to follow this convention exactly would have been impossible to attain without presenting sounds that were either uncomfortably loud or inaudibly soft. For this reason, we used the convention described above, where the 2nd note was randomly sampled from the 13th, 14th, and 15th values of the 27-step scales.

The 200 trials in each condition were presented in a different random order for each participant and dimension. The presentation order for the dimensions was determined using a Latin square design, in which one third of the participants completed the tasks in the order pitch-brightness-loudness, one third in the order loudness-pitch-brightness, and one third in the order brightness-loudness-pitch.

Predictors. Certain contour-based principles of melodic continuation have been well established and supported by previous studies of melodic continuation in pitch (Larson, 2004; Schellenberg et al., 2002; Schellenberg, 1997; Temperley, 2008). We identified three principles that had received the most empirical support from these earlier studies: Proximity, Inertia, and Post-skip Reversal.

The first predictor, Proximity, refers to the difference, in terms of scale steps, between the second and third notes, where positive values indicate that the third note was higher than the second. Previous research on pitch-based melodic expectancy has found that small absolute values of Proximity are more expected than large ones (Cuddy &

Lunney, 1995; Schellenberg et al., 2002; Temperley, 2008). Natural sound sources tend to stay within a limited pitch range, so large and rapid variations in pitch can be interpreted as the presence of multiple sound sources, which runs counter to the aim of creating the sense of a coherent melody (A. Bregman, 1990; Huron, 2001).

The second predictor, Inertia, corresponds to an expectation for pitch-based melodies to continue in the same direction after a small step (Larson, 2004). This principle can be interpreted as reflecting the Gestalt principle of good continuation (Balch, 1981), as applied to individual musical voices: once a direction has been established, a continuation of the established direction is expected.

The third predictor, Post-skip Reversal, reflects the tendency for a melody to move in the opposite direction following a large leap. This principle may reflect the tendency of melodies with good continuation, or auditory stimuli perceived as individual sound sources, to limit themselves to a restricted range of notes throughout the melody, and so to regress to the mean of that range after a leap (Temperley, 2008; von Hippel & Huron, 2000).

Among the contour-based predictors, we selected Proximity because it is one of the most broadly supported by evidence (Cuddy & Lunney, 1995; Schellenberg et al., 2002; Temperley, 2008). Post-skip Reversal is also well supported, though there is the question of whether it merely represents regression to the mean (von Hippel & Huron, 2000). There is less evidence for Inertia, with some studies, including our model study, finding no support for it (Cuddy & Lunney, 1995; Schellenberg, 1997). However, we included it in our analysis firstly because there is other evidence that supports it (Larson, 2004), and secondly because along with its symmetrical counterpart, Post-skip Reversal, it provides a general picture for which contours are expected for both small and large context intervals.

These are far from the only principles of melodic continuation that are supported by evidence, and there were alternative predictor variables we could have selected. However, many of these are disqualified from the present study because they are based on tonality, and as such there is no way to evaluate them in the dimensions of brightness and loudness. For example, one well-supported predictive principle favors continuation tones that are the tonic (primary) note of a musical key containing the previous two notes (Cuddy & Lunney, 1995). But this predictor could not be applied to brightness or loudness sequences, as musical keys cannot be formed in those dimensions. Tonality-based principles of melodic expectation, however well supported they may be, are not the concern of the present study, which seeks to compare contour-based expectations across pitch, brightness and loudness sequences.

In part, our expectations were that lower absolute values of Proximity would lead to higher ratings, and that both Post-skip Reversal and Inertia would be generally supported by our data, but our primary hypothesis was that listeners' expectations would be similar for contours in loudness and brightness to expectations for contours in pitch. Thus, the exact choice of predictors was less critical than the comparison of responses across the three auditory dimensions.

To evaluate the strength of these principles against our data, we coded each melody heard by listeners with a value indicating the degree to which that melody fulfilled each principle. Proximity was coded as the absolute difference, in steps, between the second and third notes in a melody. For example, if the second and third notes were the same, Proximity was 0, and if the third note was 12 steps down from the second note, Proximity was -12. Inertia was coded as True when a small interval (2 or 3 steps) was followed by a continuation in the same direction, False when a small interval was followed by a continuation in the opposite direction, and Neutral for any large context interval (9 or 10 steps). Post-skip Reversal was coded as True when a large interval (9 or 10 steps) was followed by a continuation in the opposite direction, False when a large interval was followed by a continuation in the same direction, and Neutral for any small context interval (2 or 3 steps). For both of these predictors, we expected true values to produce higher ratings than false values.

This produced three predictor variables, which we later compared against listener ratings. Bayesian ordinal-regression (Congdon, 2006) and repeated-measures analyses of variance (ANOVAs) were used to evaluate the significance of the contribution of each predictor in the three auditory dimensions.

Results

Figure 2.3 shows the means and between-subject standard errors of the ratings from all participants (thick solid line, n = 18), as well as means for the upper quartile (dotted line, n = 4) and lower quartile (thin solid line, n = 5) of participants ranked by musical experience. Ratings are plotted as a function of each predictor: Proximity, Inertia, and Post-skip Reversal.

As expected based on earlier studies of the perception of pitch-based melodies (Schellenberg et al., 2002; Schellenberg, 1997), ratings in the pitch dimension were highest for small absolute values of Proximity, and decreased as the size of the interval between the second and third notes increased. Our new results show that the same general pattern also holds for both brightness and loudness (Fig. 2.3, left column). These rating data were fitted using an ordinal-regression model with asymmetric Gaussian functions (Kato, Omachi, & Aso, 2002) of the predictor (Proximity). Based on 95% credible intervals (Bayesian confidence intervals, CI), the mean (μ) of the fitted Gaussians did not differ significantly from zero for any of the three dimensions: pitch: $\mu = 0.88$, CI = [-0.79; 2.70]; brightness: $\mu = -1.62$, CI = [-3.45; 0.28]; loudness $\mu = 1.59$, CI = [-0.50; 3.84]. For loudness, the difference between the upper and lower slopes of the fitted asymmetric Gaussians (Δ) was significantly larger than zero, $\Delta = 0.62$, CI = [0.34; 0.92], reflecting lower ratings with an increasingly loud final tone; for pitch and brightness, no significant asymmetry was observed, $\Delta = -0.05$, CI = [-0.28; 0.18] and $\Delta = 0.09$, CI = [-0.17; 0.39], respectively.

Although the shape of responses as a function of Proximity was very similar across the three auditory dimensions, some “fine structure” was observed in the pitch ratings that did not appear to be present in the other dimensions. For instance, dips were observed at 6 semitones, corresponding to a musical interval of an augmented fourth. This fine structure was clearer in the most musically trained listeners (dotted lines) and was not apparent in the ratings of the least musically trained listeners (thin solid lines).

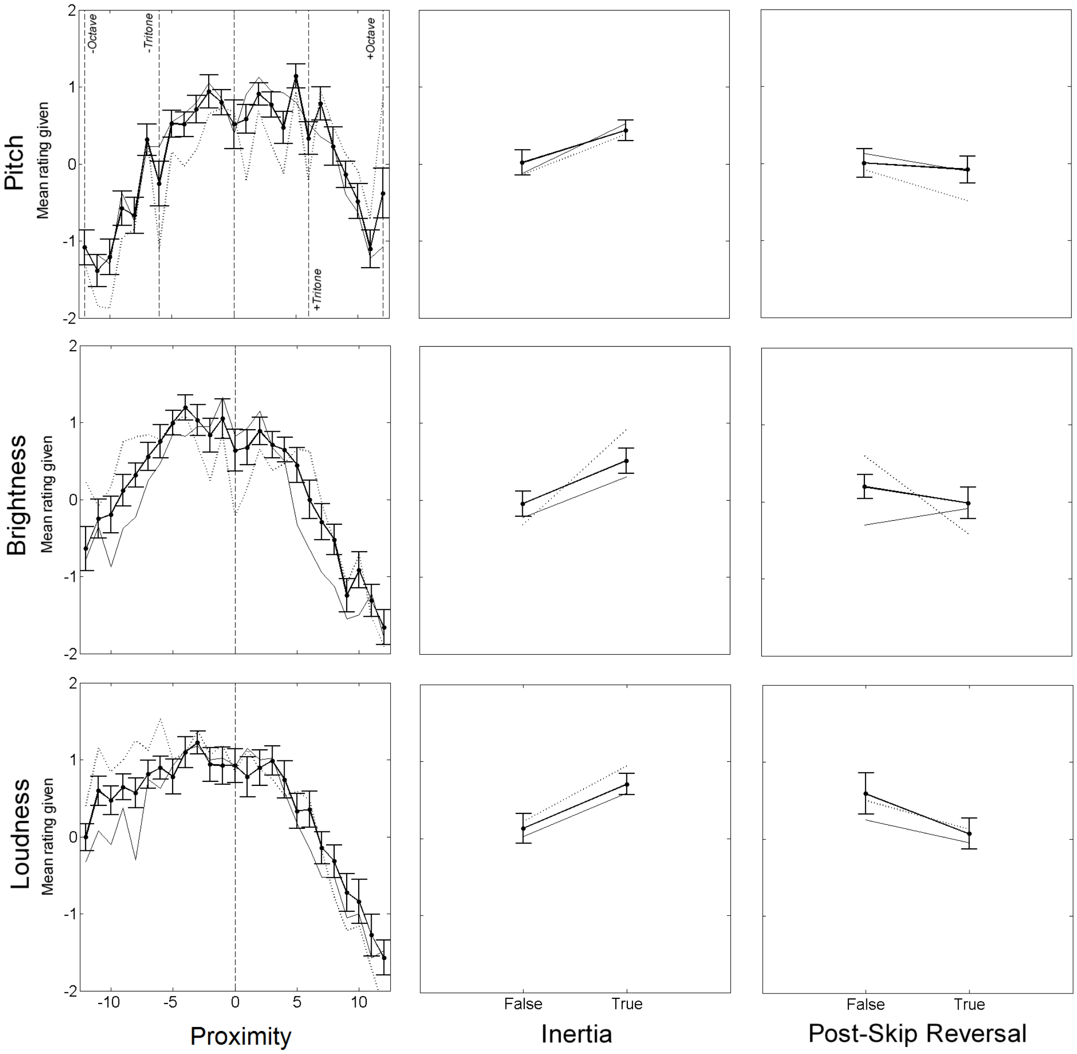


Figure 2.3. From Experiment 2.1, listener ratings of continuation for three-tone sequences in pitch (top), brightness (center), and loudness (bottom). Columns correspond to the three predictors (Proximity, Inertia, and Post-skip Reversal). Vertical dashed lines mark important values of the Proximity predictor. Mean ratings from all subjects are plotted in black with error bars +/- one standard error (between subjects). Dotted lines show mean ratings from the 4 subjects with 12 or more years of musical training. Thin solid lines show mean ratings from the five subjects with 1 or fewer years of musical training.

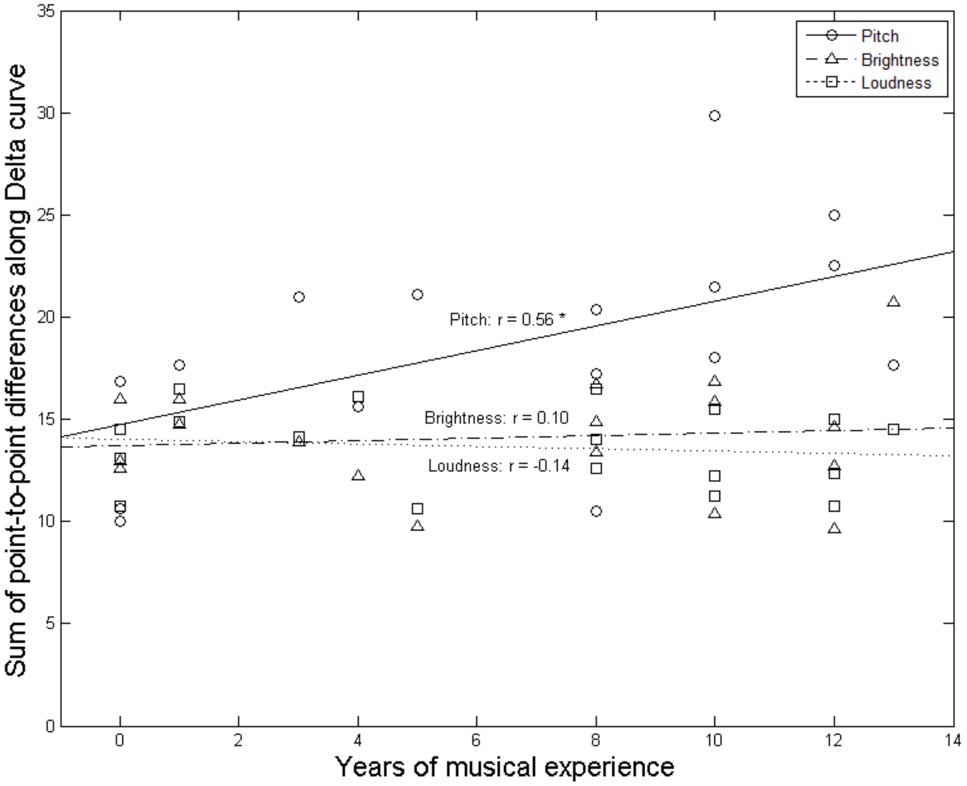


Figure 2.4. Summed absolute point-to-point differences in ratings along the Proximity curve as a function of years of musical experience. Least-squares lines are plotted for all three dimensions along with correlation coefficients I. The asterisk (*) indicates a significant correlation at p < 0.05.

In order to quantify the degree of non-monotonic fine structure or "jaggedness" in ratings along the Proximity predictor, we summed the point-to-point absolute differences in ratings along this curve for each subject in each dimension. The results are plotted in Figure 2.4, as a function of the number of years of musical training experienced by each subject. A one-way repeated-measures ANCOVA considering dimension (within subjects) and years of musical experience (between subjects) showed a significant main effect of dimension, $F(2,32) = 11.359$, $p < 0.001$, $\eta^2 = 0.415$, a significant main effect of musical experience, $F(1,16) = 9.288$, $p = .008$, $\eta^2 = 0.367$, and a significant interaction between musical experience and dimension, $F(2,32) = 3.377$, $p = .047$, $\eta^2 = 0.174$.

The results from Experiment 2.1 lend support to the expectation for fulfillment of Inertia, i.e., a melody continuing in the same direction after a small step. A two-way repeated-measures ANOVA considering dimension (pitch, brightness, or loudness) and fulfillment of Inertia (true or false; only small context intervals considered) showed a significant main effect of Inertia fulfillment, $F(1,17) = 6.23$, $p = 0.023$, $\eta^2 = 0.268$, but no significant main effect of dimension, $F(2,34) = 2.28$, $p = 0.117$, $\eta^2 = 0.119$, and no significant interaction, $F(2,34) = 0.664$, $p = 0.521$, $\eta^2 = 0.038$.

No evidence was found for a preference for Post-skip Reversal, i.e., a reversal in melodic direction after a large step: ratings either remained flat or decreased somewhat between negative and positive values of the predictor in all three auditory dimensions. A two-way repeated-measures ANOVA considering dimension (pitch, brightness, or loudness) and fulfillment of Post-skip Reversal (true or false, only large context intervals considered) showed no significant main effect of Post-skip Reversal fulfillment, $F(1,17) = 1.24$, $p = 0.282$, $\eta^2 = 0.068$. The main effect of dimension was significant, $F(2,34) = 4.9$, $p = 0.014$, $\eta^2 = 0.224$, presumably reflecting the fact that ratings for large implicative intervals were generally more positive in the loudness dimension. However, there was no significant interaction, $F(2,34) = 1.82$, $p = 0.172$, $\eta^2 = 0.098$, suggesting that the effect of Post-skip Reversal fulfillment was similar across the three auditory dimensions.

Discussion

Overall, the ratings were very similar across the three auditory dimensions, with the predictors providing similar accounts of the data. In terms of coarsely-grained expectations for broad contour, no special status for pitch was found. However, on a more fine-grained level, there were some notable non-monotonicities observed in the pitch ratings in the most musically trained listeners that were absent in the brightness and loudness ratings.

The higher ratings at Proximity values of 7 and 12 semitones in either direction correspond to musical intervals of a perfect fifth and octave, respectively, which are considered in Western musical traditions to be the most consonant intervals, whereas the lower ratings at Proximity values of 1, 6, and 11 semitones, correspond to musical intervals of a minor second, augmented fourth, and major seventh, which are considered to be among the most dissonant (McDermott, Lehr, & Oxenham, 2010; Plomp & Levelt, 1965). The fact that preference for tonal consonance is stronger in musically trained listeners is consistent with many earlier findings (Krumhansl & Shepard, 1979; McDermott, Lehr, et al., 2010), and is consistent with the prevailing view that these preferences may be learned through training and exposure (Szpunar, Schellenberg, & Pliner, 2004; Thompson et al., 2000). On the other hand, the observed interaction between musical experience and "jaggedness" of ratings along the Proximity predictor may simply reflect increased sensitivity to pitch changes in musically trained listeners, in the absence of increased sensitivity to brightness and loudness changes. This is an empirical question that could be resolved with future research.

The absolute interval size (Proximity) was a strongly supported predictor, with smaller absolute values predicting higher ratings. In this respect our results are consistent with converging evidence for contour-based principles of melodic expectancy from two experimental paradigms: subjective ratings of continuation (Cuddy & Lunney, 1995;

Schellenberg et al., 2002; Schellenberg, 1996; Schmuckler, 1989) and production, by singing or playing, of the note considered most likely to follow a melodic context (Carlsen, 1981; Schellenberg et al., 2002).

The results of Experiment 2.1 also lend support to the principle of Inertia, with fulfillment of this principle linked to higher ratings across melodies in all three stimulus dimensions. This is consistent with some previous support for this principle (Larson, 2004), but inconsistent with other studies that found no evidence for it (Cuddy & Lunney, 1995; Schellenberg, 1997).

Post-skip Reversal was not supported in any dimension, which seems contrary to both our expectations and the existing evidence. However, Post-skip Reversal may in fact be an emergent property, reflecting the restricted range of most melodies (von Hippel & Huron, 2000). Indeed, explicit prescriptions for small intervals between notes and for narrow overall ranges, taken together, produce an expectation for a small step in the opposite direction following a large leap, or a regression towards the mean (Temperley, 2008; von Hippel & Huron, 2000).

The explanation of Post-skip Reversal in terms of regression towards the mean may account for why it was not a strong predictor in Experiment 2.1. In the present study, individual trials occurred in quick succession and it is likely that listeners retained some memory of recent trials, making it plausible that subjects were basing judgments in part relative to the overall range of stimuli presented in the experiment. The second note in our paradigm was always taken from the middle of range of possible notes in the scale (the 13th, 14th, or 15th member of the 27-step scale). Therefore, a large context interval called for the first note, not the second note, to fall at an extreme end of the range. The first note in a large context interval was especially likely to sound “extreme” in Experiment 2.1 because the context interval changed from trial to trial, so it is likely that context intervals in the immediately preceding trials were small, or went in the opposite direction, or both. In this way, Experiment 2.1 effectively dissociated Post-skip Reversal from a regression towards the mean, and the results may imply that, once dissociated, Post-skip Reversal may not play an important role in predicting expectations. However, this conclusion may only be valid for short melodies such as we used in our experiments, reflecting a general expectation for continuation in any short sequence. It remains possible that longer melodies may still produce expectations for Post-skip Reversal, even when the skip does not end on a value far from the mean of recently heard notes.

The question remains why we did not find Post-skip Reversal to be a significant predictor, whereas Cuddy and Lunney (1995) did, while using a very similar paradigm. One important difference may be our randomized presentation of context intervals, compared with their blocked presentation method, which resulted in the same context interval being presented 25 times in a row. The other difference was that they alternated the second note of this context interval between C4 and F#4 on every trial, whereas in Experiment 2.1 the second note was randomly selected from a set of three intermediate values.

Eliminating both of these paradigmatic differences and replicating the experiment of Cuddy and Lunney as exactly as possible may produce similar results to theirs in the pitch dimension. Presenting the same context interval 25 times in a row shifts the overall mean of recently heard tones towards the mean of that context interval, which could cause listeners to expect regression towards that mean in the form of Post-skip Reversal, “filling in” the context interval. This is only true if the same absolute pitches are used on every trial, but that condition is almost fulfilled by alternating between C4 and F#4. It seems plausible that listeners could form templates based on alternating trials and that some form of “build up” could occur. Experiment 2.2 was designed to test this possibility by replicating the design of Cuddy and Lunney (1995) as closely as possible, and by extending their paradigm to the dimensions of brightness and loudness.

III. Experiment 2.2: Repeated melodic context

Rationale

The results from Experiment 1 supported our initial hypothesis that contour-based melodic expectation generalizes to auditory dimensions other than pitch. One aspect of the data, however, was not consistent with an earlier study of melodic expectation. In contrast to the results of Cuddy and Lunney (1995), we found no significant effect of Post-skip Reversal in any dimension, whereas they had found an effect using pitch contours. We ascribed this difference to their use of stimuli that were blocked by context interval. The current experiment had two main aims. The first aim was to attempt to replicate the findings of Cuddy and Lunney (1995) by using stimuli that were blocked by context interval. The second aim was to compare the responses in this altered paradigm across the three auditory dimensions tested in Experiment 2.1. If changes in the stimulus presentation method led to similar changes in all three dimensions, the results would further support our main hypothesis that contour-based melodic expectations generalize beyond the dimension of pitch.

Method

Stimuli. Harmonic complex tones were generated in the same way as Experiment 2.1. The "scales" created in Experiment 2.1 were adjusted slightly to increase the number of available steps from 27 to 33. In pitch, F0s ranged from C3 (131 Hz) to F#5 (741 Hz) in 1-semitone steps, with the 2nd note of every melody alternating between C4 (262 Hz) and F#4 (370.5 Hz). In brightness, the center frequency of the Gaussian function ranged from 131 Hz to 4192 Hz, in 2-semitone steps, with the 2nd note alternating between 524-Hz and 1048-Hz centroids. In loudness, the step size had to be decreased to 1.5 dB (down from 2 dB in Experiment 1), to allow for 33 levels ranging from 30 to 79.5 dB, with the 2nd note alternately 48 or 57 dB. As in Experiment 1, when the stimuli varied along a single dimension, the others were held constant. The constant values for the three dimensions were 131 Hz for F0, 800 Hz for spectral center frequency, and 50 dB SPL for sound level.

The same 8 context intervals (±2, ±3, ±9, and ±10 steps) and 25 continuation tones (12 steps below to 12 steps above the second tone) were tested, and the 1200ms-400ms-800ms pattern in stimulus onset asynchronies was also retained.

Procedure. Eighteen new listeners, 3 male and 15 female, were recruited from the Twin Cities campus of the University of Minnesota, ranging in age from 18 to 39 (*M* = 23.8, *SD* = 5.1). This group of participants reported an average of 6.2 years of musical training (*SD* = 7.2; range 0 to 20 years). The five participants who reported no musical training and the five participants who reported the highest amount of musical training (at least 8 years) were taken as an approximation of the lower and upper quartiles, respectively, of participants ranked by musical experience. Once again, all listeners had normal audiometric hearing thresholds (defined as not exceeding 20 dB HL for octave frequencies between 250 and 8000 Hz).

Listeners heard and rated melodies in the same way as Experiment 2.1, with two important exceptions. First, as noted above, the 2nd note of the melody alternated from trial to trial between the 13th and 19th notes of the 33-step scale. Second, the melodies were blocked by context interval, such that melodies beginning with the same context interval were all presented in immediate succession instead of being randomized from trial to trial. The presentation order for the dimensions was again counterbalanced with a Latin Square design.

Results

Figure 2.5 shows the means and between-subject standard errors of the ratings from all participants in Experiment 2.2 (thick solid line, n = 18), as well as means for the upper quartile (dotted line, n = 5) and lower quartile (thin solid line, n = 5) of participants ranked by musical experience. Ratings are plotted as a function of each predictor: Proximity, Inertia, and Post-skip Reversal.

The pattern of results along the proximity predictor was broadly similar to the pattern found in Experiment 2.1. Once again, small absolute values of Proximity produced higher ratings in all three dimensions (Fig. 2.5, left column). We applied the ordinal-regression model introduced in Experiment 1, with asymmetric Gaussian functions of Proximity fitted to the data. The 95% Bayesian confidence intervals (CI) identified by this analysis found no evidence that the mean (μ) of the fitted Gaussians differed from zero for any of the three dimensions: pitch: $\mu = 0.95$, CI = [-3.27 4.31]; brightness: $\mu = -0.98$, CI = [-3.22 1.33]; loudness $\mu = 1.98$, CI = [-0.27 4.38]. The difference between the upper and lower slopes (Δ) of the fitted Gaussians was significantly larger than zero only in the loudness dimension, Δ = 0.36, CI = [0.18; 0.55], again reflecting lower ratings with an increasingly loud final tone; this asymmetry was not significant in pitch, Δ = -0.13, CI = [-0.41 0.11], or brightness, Δ = -0.03, CI = [-0.13; 0.22]. Also similarly to Experiment 2.1, there appears to be more fine-grained non-monotonicity in the ratings for pitch than in the other two dimensions, with characteristic dips at the tritones, and this effect appears more pronounced among the most musically trained listeners.

The results from Experiment 2.2 provided no support for the expectation for fulfillment of Inertia, i.e., a melody continuing in the same direction after a small step. A two-way repeated-measures ANOVA considering dimension (pitch, brightness, or loudness) and fulfillment of Inertia (true or false; only small context intervals considered) found neither a significant main effect of fulfillment, $F(1,17) = 0.048$, $p = 0.829$, nor a main effect of dimension, $F(2,34) = 0.063$, $p = 0.939$, and no interaction, $F(2,34) = 0.070$, $p = 0.932$.

In contrast, the results provided significant support for Post-skip Reversal, i.e., a reversal in melodic direction after a large step. A two-way repeated-measures ANOVA considering dimension (pitch, brightness, or loudness) and fulfillment of Post-skip Reversal (true or false, only large context intervals considered) showed a significant main effect of fulfillment, $F(1,17) = 5.38$, $p = 0.033$, $\eta^2 = 0.241$, but no significant main effect of dimension, $F(2,34) = 1.819$, $p = 0.178$, and no significant interaction, $F(2,34) = 0.649$, $p = 0.529$.

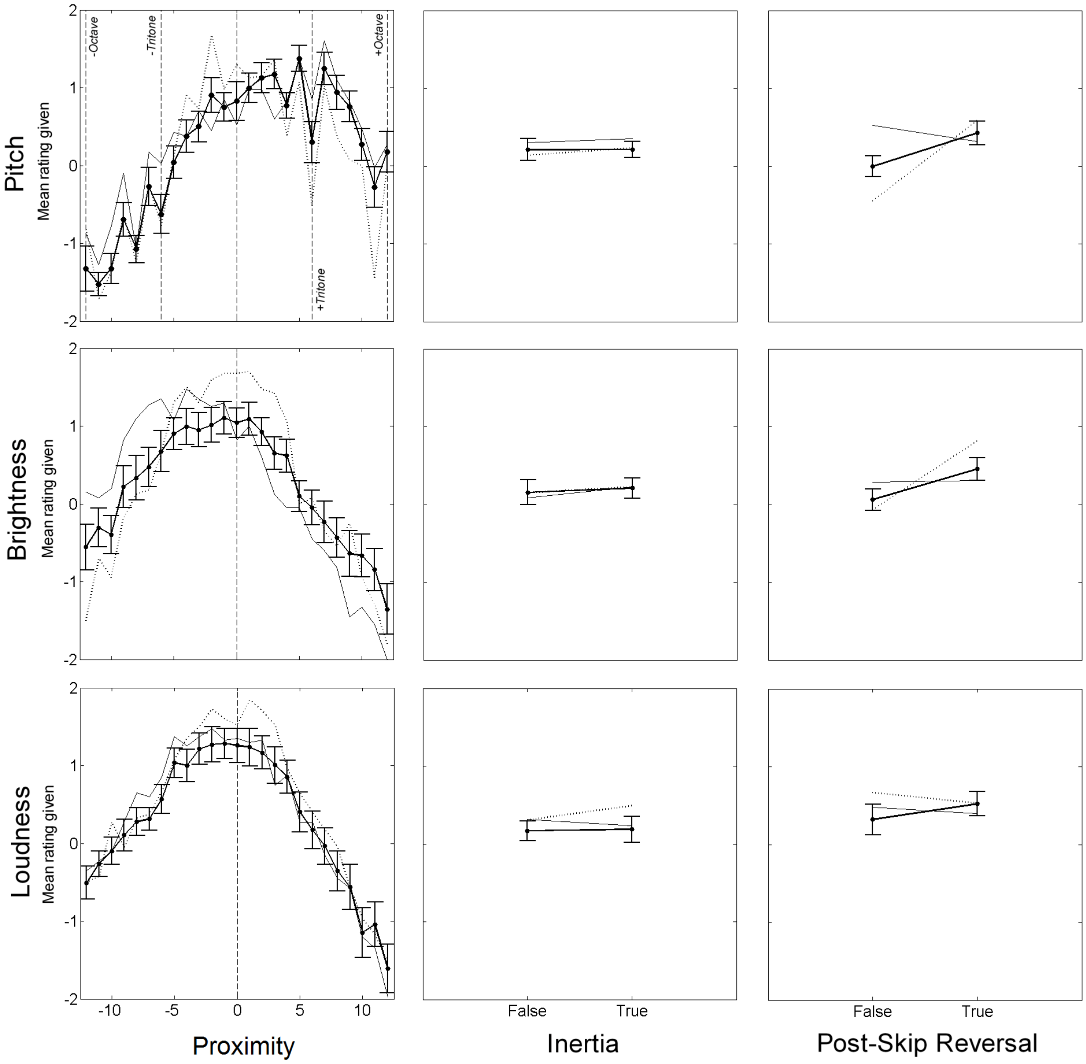


Figure 2.5. From Experiment 2.2, listener ratings of continuation for three-tone sequences in pitch (top), brightness (center), and loudness (bottom). Columns correspond to the three predictors (Proximity, Inertia, and Post-skip Reversal). Vertical dashed lines mark important values of the Proximity predictor. Mean ratings from all subjects are plotted in black with error bars +/- one standard error (between subjects). Dotted lines show mean ratings from the 5 subjects with 8 or more years of musical training. Thin solid lines show mean ratings from the 5 subjects with 0 years of musical training.

Discussion

As predicted, we successfully replicated Cuddy and Lunney's (1995) results in pitch by more precisely matching their paradigm in presentation order and absolute pitch selection. Experiment 1 found no support for Post-Skip Reversal, when context intervals were presented in random order from trial to trial, but in Experiment 2.2, when the melodies were blocked by context interval, the ratings lent support to the principle. More importantly, this substantive change in the pattern of results was observed in all three dimensions.

Taken together, the results from our two experiments, along with those of previous studies (Cuddy & Lunney, 1995; Schellenberg et al., 2002), suggest that some properties of melodic expectation, such as Post-skip Reversal and Inertia may be critically dependent on the presentation method. It could be argued that the randomized presentation method of Experiment 2.1 is more valid than the blocked method of Experiment 2.2, and that Post-skip Reversal in particular may simply reflect a more general principle of tendency towards the mean (Temperley, 2008; von Hippel & Huron, 2000). However, the question of whether certain predictors are valid is tangential to the main finding of the present study: regardless of the predictors and the methods used, the results from pitch, brightness, and loudness remain similar. This outcome further supports the hypothesis that contour-based expectations for melodic continuation generalize beyond the auditory dimension of pitch.

IV. General discussion

The purpose of our study was to test whether certain broadly supported principles and features of contour-based expectations for melodic continuation are specific to pitch, or whether they generalize to other auditory dimensions. We found substantial agreement between the ratings for sequences in all three auditory dimensions and established that the predictors that were successful in predicting expectations in pitch were similarly successful in the dimensions of brightness and loudness.

Composers such as Arnold Schoenberg and Anton Webern have composed melodic contours in timbre by switching melodies rapidly from instrument to instrument, a technique called *Klangfarbenmelodie* (Schoenberg, 1911). The present study found that listener expectations for musical contour can be fulfilled or violated not only by changes in pitch, but also by changes in timbre or loudness. This finding provides empirical evidence for the validity of composing melodic contours in dimensions other than pitch.

Melodic expectancies are stimulus-stimulus expectancies, where one stimulus type implies another stimulus type. This specific kind of expectancy is an essential part of learning (Bolles, 1972). More broadly, expectancies, as a general cognitive phenomenon, play a large part in determining behavior (Kirsch, 1985). Expectancies are often studied specifically through perception of musical melodies in pitch, as a controlled and limited context from which more general conclusions concerning expectancies can be drawn (Dowling, 1990; Schellenberg et al., 2002). Although it explores only auditory perception, the present study provides some evidence for the previously unsupported assumption that patterns of expectation for melodies can be generalized beyond the context of musical melodies defined by changes in pitch.

Overall, the results suggest that the principles of good melodic continuation, described in many earlier studies of both experimental psychology and music theory, are not specific to melodies, as traditionally defined by pitch movement. Instead they may reflect general principles that extend to many auditory dimensions. Specifically, the principles involving interval size (Proximity) and trajectory (Inertia) may be viewed as expressions of basic principles of auditory perceptual organization: sequential sounds that vary by only a small amount in any given dimension, and continue within a limited trajectory, are more likely to form a single “auditory stream.” Thus, as suggested by Huron (2001), expectations for melodic continuation and voice leading may reflect principles that encourage perceptual binding. Our results extend and generalize this conclusion to perceptual dimensions other than pitch, and suggest that rules of melodic continuation have not emerged from exposure to specific music or pitch-based melodies, but may instead reflect fundamental principles of perceptual organization that transcend the specific dimension of pitch.

Chapter 3: Melodic contour perception of listeners with congenital amusia

I. Introduction to congenital amusia and melody perception

Congenital amusia is a disorder characterized by specific deficits in music perception, despite otherwise normal auditory capabilities for non-musical stimuli (Ayotte, Peretz, & Hyde, 2002). One explanation for the specificity of this disorder to music is that it involves deficits in fine-grained pitch perception, whereas timing and rhythm perception is intact (Hyde & Peretz, 2004), and responses to speech stimuli are normal (Tillmann, Schulze, & Foxton, 2009). Although recent findings (Whiteford & Oxenham, 2017) have raised some doubts about this canonical explanation, the hypothesized connection between amusia and pitch-specific auditory mechanisms makes amusia a topic of interest for pitch researchers, allowing for insight into phenomena like consonance (Cousineau, McDermott, & Peretz, 2012) and harmonic resolvability (Cousineau, Oxenham, & Peretz, 2015). This interest is reciprocal: a better understanding of amusia will also require a better understanding of pitch perception mechanisms in general.

Pitch, a perceptual correlate of the overall periodicity of a sound, is the auditory dimension that allows us to perceive melody and harmony in music, prosody in speech, and even semantic meaning in tone languages. Pitch contour, or the relative pattern of changing pitch directions over time, is encoded in memory to allow for recall of melodies – both in short-term memory (STM), allowing for immediate comparison with new stimuli, and in long-term memory (LTM), allowing for future recognition and reproduction. Contour is not the only aspect of melody stored in memory: the melody's tonality, or position within a scale or hierarchy, is also encoded (Dowling, 1978). Some coarse information about the absolute pitches of the melody appears to persist in LTM (Levitin, 1994), but from an early age, humans encode melodies into memory mainly in terms of relative pitch (Plantinga & Trainor, 2005). Newly learned melodies quickly become robust to alterations of absolute parameters such as key and tempo (Schellenberg & Habashi, 2015), suggesting that they are mostly stored as relative pitch information, including pitch contour.

The equal availability of other auditory cues has led some researchers to question why pitch is so central as a cue for recognition and recall in music. In fact, listeners are also capable of recognizing melodic contours in at least two other auditory dimensions, brightness (an aspect of timbre) and loudness (McDermott et al., 2008). These contours in brightness and loudness can be reliably compared within or across dimensions, and also support recognition of familiar memories from LTM. They even produce melodic expectations similar to those found in pitch melodies, i.e. expectations for small intervals and regression towards the mean pitch (Graves, Micheyl, & Oxenham, 2014). So why are they not used in the same way as pitch in musical compositions? It may be that pitch is more reliably coded than other dimensions – relative to interval discrimination thresholds, basic difference limens for pitch are low compared to other dimensions (McDermott, Keebler, et al., 2010). But even when stimuli are equated for discriminability, pitch seems to have an advantage over loudness for processing of sequences (Cousineau, Demany, & Pressnitzer, 2009). The centrality of pitch is not universal across species; for instance, songbirds appear to use a spectral shape cue instead of pitch to recognize melodies (Bregman, Patel, & Gentner, 2016). In any case, many studies have shown that pitch, brightness, and loudness are not completely independent of each other – the perceived pitch of a tone is influenced by changes in intensity as well as changes in brightness (Allen & Oxenham, 2014; Krumhansl & Iverson, 1992; Melara & Marks, 1990; Pitt, 1994; Warrier & Zatorre, 2002).

Whatever sets pitch apart from other dimensions in music may also explain why its perception is specifically impaired in congenital amusia. Amusia has long been reported to involve difficulty recognizing familiar melodies and perceiving melodic contours (Ayotte et al., 2002), though some evidence suggests that amusics might implicitly store familiar melodies in LTM, lacking explicit access to identify the melodies, but able to distinguish novel from previously heard melodies (Tillmann, Albouy, Caclin, & Bigand, 2014). Since it is still not clear to what extent amusia is a pitch-specific disorder, our goal was to determine whether the amusia-related deficit in melodic contour perception extends to auditory dimensions other than pitch.

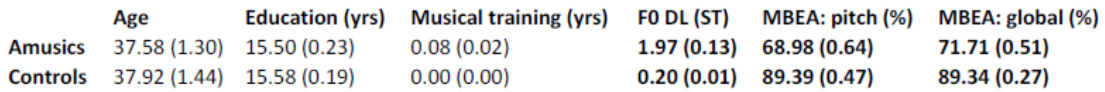
Some recent evidence (Lu, Sun, Ho, & Thompson, 2017) suggests that in congenital amusia, access to auditory contour cues may be less impaired for brightness and loudness than it is for pitch, at least in the context of cross-modal integration with a visual cue. It remains unclear, however, whether amusics exhibit deficits in intra-modal contour perception within the auditory dimensions of brightness and loudness. As a secondary question, we were also interested in the role tonality may play in melody recognition for amusics, given that they may be implicitly sensitive to tonal structure as measured by reaction time (Albouy, Schulze, Caclin, & Tillmann, 2013). Since brightness and loudness melodies lack tonality, they provide a novel way to evaluate the effect of tonality on melody perception.

In order to test the degree to which the contour processing deficit associated with amusia is specific to pitch, we conducted two experiments, the first primarily involving LTM and the second primarily involving STM. In the first experiment, listeners were asked to identify familiar melodies when defined by varying pitch intervals, stretched pitch intervals, brightness, loudness, or just rhythm. In the second experiment, listeners were asked to compare the contours of novel melodies when defined in terms of pitch, brightness, or loudness contours. In both experiments, we evaluated the size of the amusia-related deficit as a function of the auditory dimension in which the stimuli were presented.

II. General method for Experiments 3.1 and 3.2.

Listeners

We recruited 12 amusic individuals (7 female and 5 male) and 12 yoked control participants (7 female and 5 male), matched in terms of age, education, musical experience, gender, and handedness. Descriptions of the two groups are provided in Table 3.1. All participants were screened for normal audiometric hearing thresholds, and congenital amusia was identified using the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz, Champod, & Hyde, 2003). All participants provided written informed consent and were compensated for their participation.



	Age	Education (yrs)	Musical training (yrs)	F0 DL (ST)	MBEA: pitch (%)	MBEA: global (%)
Amusics	37.58 (1.30)	15.50 (0.23)	0.08 (0.02)	**1.97 (0.13)**	**68.98 (0.64)**	**71.71 (0.51)**
Controls	37.92 (1.44)	15.58 (0.19)	0.00 (0.00)	**0.20 (0.01)**	**89.39 (0.47)**	**89.34 (0.27)**

Table 3.1. Demographic information, F0 difference limens, and MBEA scores for 24 listeners (12 amusics and 12 matched controls). Mean values are shown, with standard deviations in parentheses. Bold values indicate independent t-tests with *p* values less than .05.

Stimuli and Procedure

In both experiments, listeners heard melodies made up of harmonic complex tones with Gaussian spectral envelopes, after McDermott, Lehr, and Oxenham (2008). The spectral envelope of each tone was defined by a Gaussian function in the linear frequency domain, with the standard deviation equal to 25% of the mean (and peak) frequency. In the present study, we also used harmonic complex tones to carry melodies in the loudness condition, as opposed to the noise bursts used in McDermott et al. (2008). Each tone was gated on and off with raised-cosine ramps that were 100 ms in Experiment 3.2, to match the ramp durations in McDermott et al. (2008), but only 10 ms in Experiment 3.1, so that on and off ramps did not overlap for very short notes in melodies.

The melodic contour was carried by one of three dimensions. When melodic contour was carried by pitch, the F0 was varied along a range from 150 to 476 Hz. When contour was carried by a different variable, the F0 stayed constant at 256 Hz. When contour was carried by brightness, the mean of the Gaussian spectral envelope was varied along a range from 315 to 3175 Hz; otherwise it was constant at 1000 Hz. When contour was carried by loudness, the sound level of the complex was varied along a range from 20 to 80 dB SPL; otherwise it was constant at 65 dB SPL. A change of 1 semitone in F0 in the pitch condition corresponded to a change of 2 semitones in spectral centroid in the brightness condition, and a change of 2 dB SPL (Exp. 3.1) or 3 dB (Exp. 3.2) in the loudness condition. The smaller step size in Experiment 3.1 is due to the wide range of certain melodies used in that experiment, and the narrow range of sound levels available.

All listeners completed Experiment 3.1 first, followed by Experiment 3.2. Some participants completed both experiments in a single session, while others completed them on separate days.

III. Experiment 3.1: Long-term memory: familiar melody recognition

Method

We devised a version of the familiar melody recognition task used by McDermott et al. (2008), modified so as to decrease the overall difficulty, in the hopes of eliciting above-chance performance in the amusic group. Instead of open-set recognition, we allowed listeners to familiarize themselves before testing with a closed set of 10 familiar French melodies, and then to respond during testing by selecting the name of the correct melody from a list of 10 possibilities. Although amusics have been reported to have difficulty recognizing familiar melodies (Ayotte et al., 2002), they also may report an implicit feeling of familiarity or recognition even when unable to name the song (Tillmann et al., 2014). We therefore also measured subjective self-rated confidence as well as objective accuracy on this task. Since some aspects of music perception in amusia may be implicit, resulting in effects on response time (RT) in the absence of effects on accuracy, we measured RT (Albouy et al., 2013).

Stimuli. From a set of melodies previously identified as highly familiar to French university students (Ehrlé, Samson, & Peretz, 2001), and inspired by another study on familiar melody recognition (Devergie, Grimault, Tillmann, & Berthommier, 2010), we chose ten melodies familiar to French audiences, displayed in Figure 3.1 in the original pitch condition. The melodies were generated in one of five conditions: pitch (no change), stretched pitch (the same melody, but with all interval sizes doubled), brightness, loudness, or rhythm only. These are a subset of the conditions used for this task by McDermott et al. (2008). In order to minimize the salience of surface-level cues that would allow the listener to perform the task without truly recognizing the melody, the melodies were equated for tempo and average pitch. Each melody’s duration from start to finish was 9.6 seconds, usually consisting of 4 measures of 4/4 time at a tempo of 100 bpm. A slight exception is *Fais dodo, Colas, mon p’tit frère*, the only melody subdivided in three, where the time signature is 6/8, meaning the tempo was 150 bpm for an eighth note or 50 bpm for a dotted quarter note. In the pitch and stretched pitch conditions, the absolute pitch of each melody was defined such that the average F0 of the melody was equal to 262 Hz (C4), meaning that the melodies were in different keys from each other, and not tuned to a consistent standard (the notation in Figure 3.1 is the closest chromatic approximation). A similar procedure was used for the brightness and loudness conditions, such that the average spectral centroid of all brightness melodies was 1000 Hz, and the average level of all loudness melodies was 65 dB SPL.

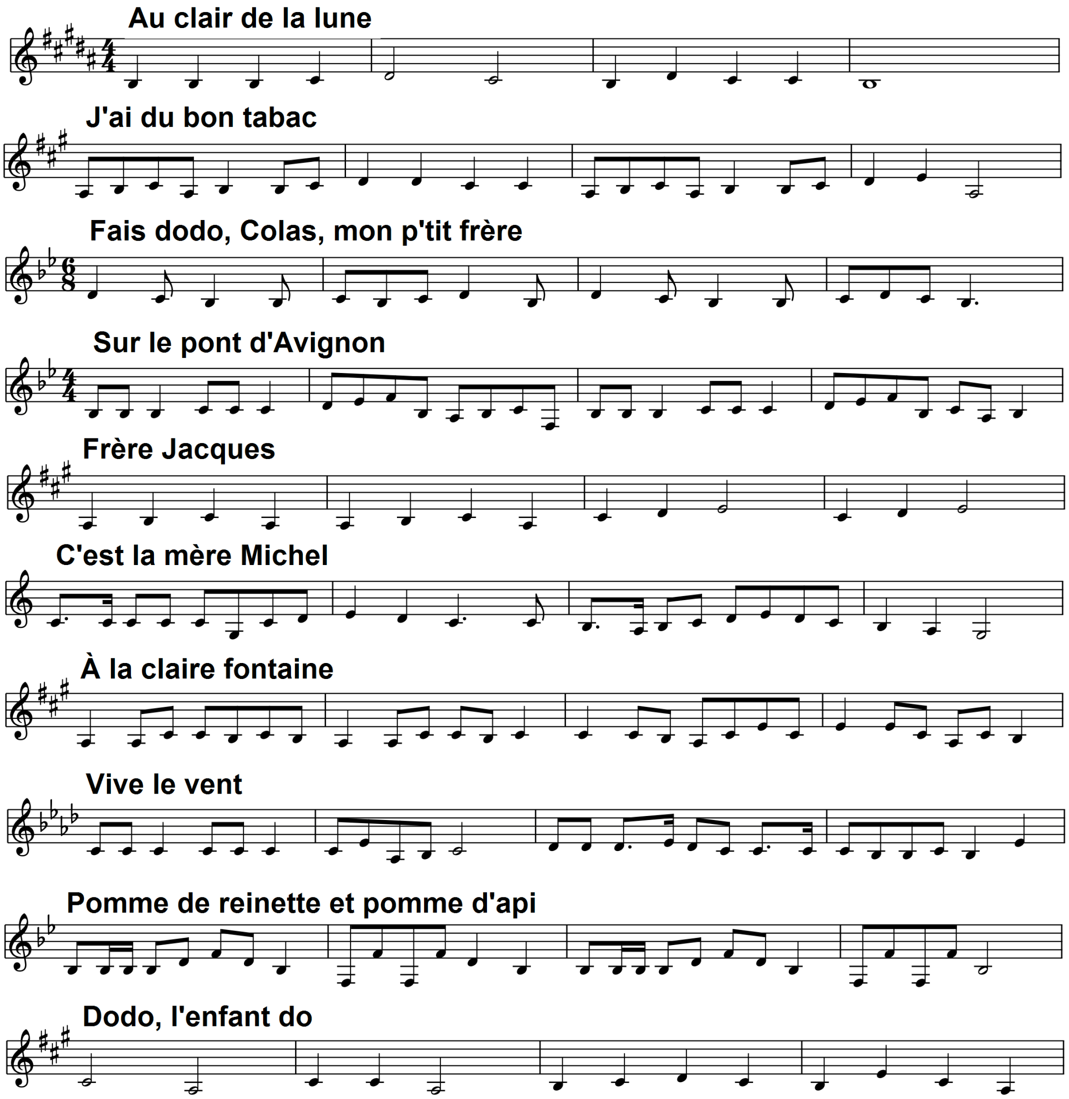
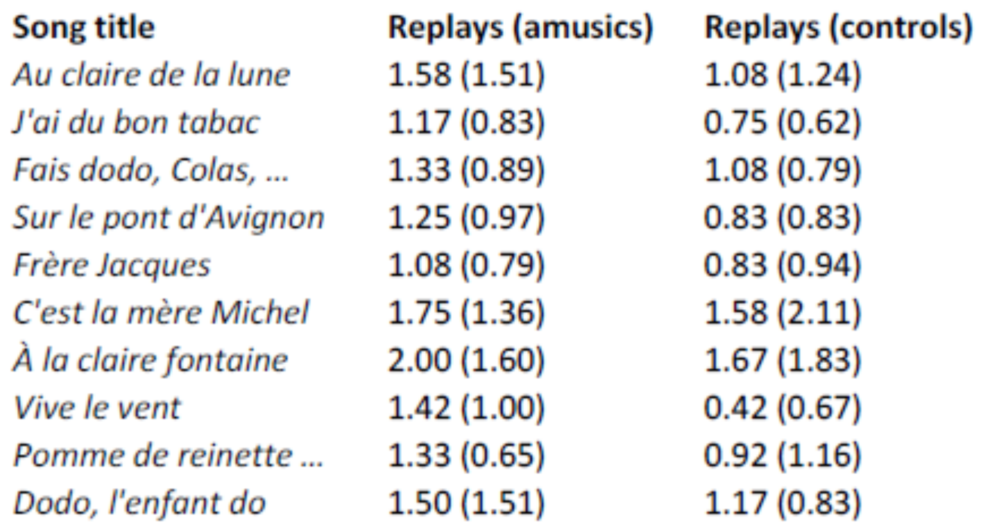


Figure 3.1. Musical notation showing the 10 familiar French melodies used for the closed-set recognition task in Experiment 3.1. All melodies last 9.6 seconds. The pitch condition is shown here. Listeners were instructed to identify these melodies when defined by pitch (as shown), stretched pitch (interval sizes doubled), brightness, loudness, or rhythm only.

Procedure. At the very beginning of the experiment, an example melody (“*Ah, vous dirai-je, maman”*) was played for the listener in all five conditions, in the following order: pitch, stretched pitch, brightness, loudness, rhythm only. The conditions were not described in detail, but the listener was encouraged to focus on the similarities between the five different versions of the melody, and informed that later in the experiment their task would be to identify melodies played in any of these five styles. After this opening orientation, the names of the ten melodies appeared in a circle on screen, and the listener was asked to listen to each melody in the original "pitch" condition at least once, by clicking on the button next to its name. Then the listener was allowed to replay each melody as many times as desired, in order to ensure familiarity and an ability to connect the name of the song with its melody. Table 3.2 summarizes the number of times amusics and controls chose to re-play each melody during this training phase.



Song title	**Replays (amusics)**	**Replays (controls)**
Au claire de la lune	1.58 (1.51)	1.08 (1.24)
J'ai du bon tabac	1.17 (0.83)	0.75 (0.62)
Fais dodo, Colas, ...	1.33 (0.89)	1.08 (0.79)
Sur le pont d'Avignon	1.25 (0.97)	0.83 (0.83)
Frère Jacques	1.08 (0.79)	0.83 (0.94)
C'est la mère Michel	1.75 (1.36)	1.58 (2.11)
À la claire fontaine	2.00 (1.60)	1.67 (1.83)
Vive le vent	1.42 (1.00)	0.42 (0.67)
Pomme de reinette ...	1.33 (0.65)	0.92 (1.16)
Dodo, l'enfant do	1.50 (1.51)	1.17 (0.83)

Table 3.2. The number of times each melody was voluntarily replayed in the training phase of Experiment 3.1, for amusic and control groups. Mean values are shown, with standard deviations in parentheses.

The testing phase, which immediately followed the training phase, consisted of 50 trials. Each of the 10 melodies was presented once in each of the 5 conditions, in a pseudorandom order. The names of the 10 melodies remained on screen throughout the experiment, but while a melody was playing, their associated response buttons were grayed out. Participants were instructed to respond as quickly as possible once the melody had finished playing, by moving the mouse and clicking on their choice. No feedback was given. After each response, a 5-point Likert scale appeared on which the listeners were to rate their level of confidence from 1 (“not at all confident”) to 5 (“very confident”).

Analysis. Three dependent variables were recorded on each trial: accuracy (0 or 1), RT (in ms), and confidence (an ordinal rating between 1 and 5). We modeled the variance in each of these 3 dependent variables using a generalized linear mixed model (GLMM) in SPSS, version 24 (“IBM SPSS Statistics for Windows, Version 24.0.,” 2016), with fixed factors of Group (amusic or control), Condition, Trial Number, and Melody, and a random factor of Subject. We also included the potential two-way interaction between Group and Condition, but no other two-way or higher-level interactions. For accuracy (0 or 1), the response distribution was binomial, with a logit link function. For RT, the response distribution was a gamma distribution, with a log link function. For confidence, the response distribution was a multinomial distribution, with a cumulative logit link function. Before analyzing RT data, outliers were excluded, pooling across subjects. An outlier was defined as exceeding the mean plus twice the standard deviation of log-transformed response times (this criterion value was 5968 ms). Out of 1200 measured RTs, 70 were excluded on this basis.

Results

Figure 3.2A shows results for accuracy by group and condition. Using a generalized linear mixed model on a binomial distribution for accuracy, we found fixed effects of Group, *F*(1,1181) = 10.49, *p* = .001, Condition, *F*(4,1181) = 40.82, *p* < .001, and Melody, *F*(9,1181) = 3.92, *p* < .001, but no effect of Trial Number and no Group by Condition interaction. For Condition, pairwise comparisons revealed that every condition produced better accuracy than Rhythm Only, with no other comparison reaching significance. For Melody, despite the overall effect, no individual melody deviated significantly from the mean of all ten melodies. Figure 3.3A shows results for accuracy by melody.

Figure 3.2B shows results for confidence by group and condition. Using a generalized linear mixed model on a multinomial distribution for confidence, we found fixed effects of Group, *F*(1,1177) = 11.99, *p* = .001, Condition, *F*(4,1177) = 98.29, *p* < .001, and Melody, *F*(9,1177) = 10.81, *p* < .001, but no effect of Trial Number. For confidence, we found a significant Group by Condition interaction, *F*(4,1177) = 3.36, *p* = .01. For Condition, pairwise comparisons revealed that every condition significantly differed from every other condition. Investigating the interaction, controls were significantly more confident than amusics for the stretched pitch, brightness, and loudness conditions, but not for pitch or rhythm only. For Melody, the following melodies produced significantly lower confidence than the average: *Au clair de la lune, Frère Jacques, À la claire fontaine,* and *Dodo, l'enfant do*. Melodies that produced higher than average confidence were *Vive le vent* and *Pomme de reinette et pomme d'api.* Figure 3.3B shows results for confidence by melody.

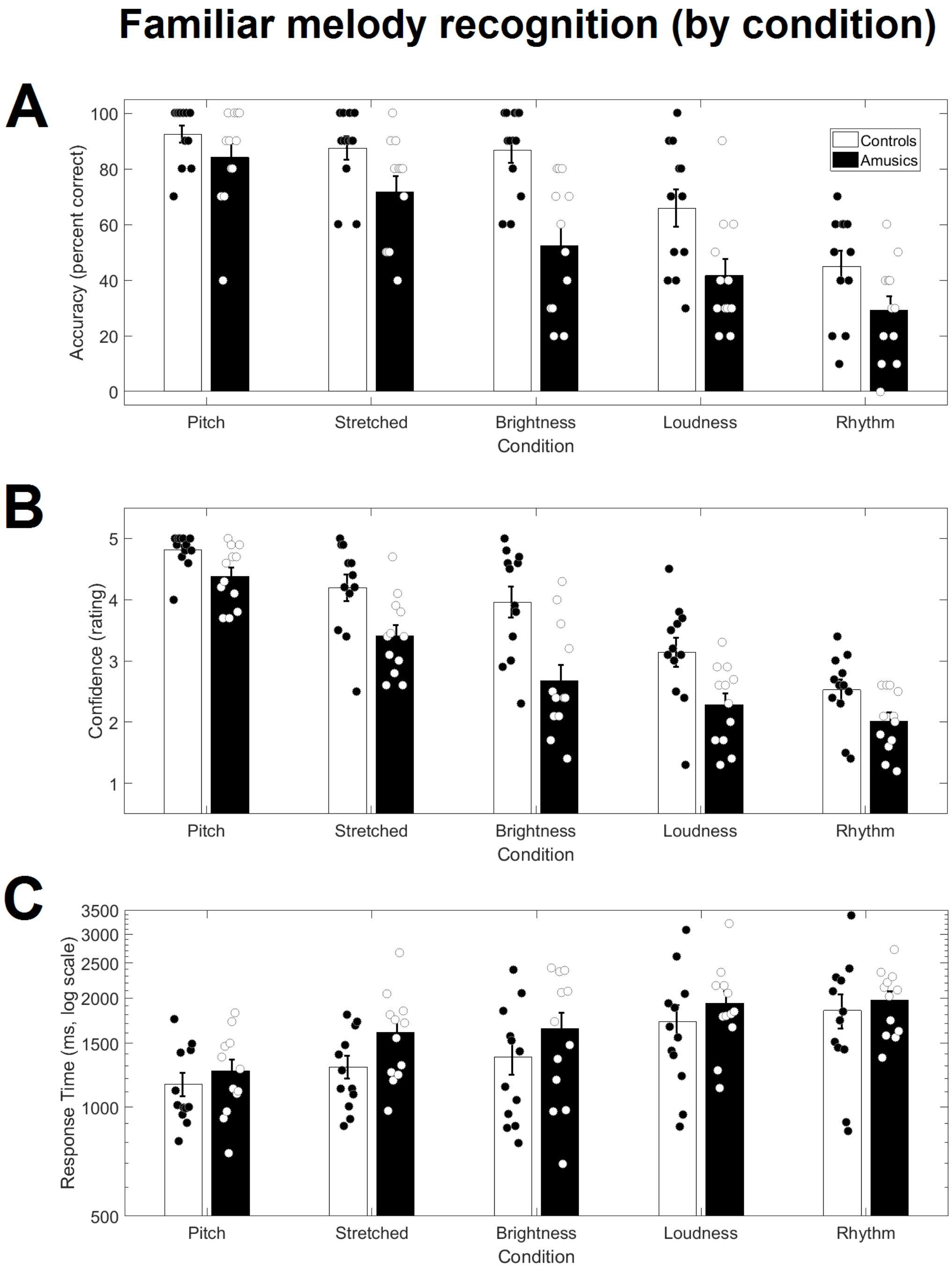


Figure 3.2. Results from Experiment 3.1, by group and condition, for accuracy (top), confidence (center), and RT (bottom). Error bars +/- 1 SEM, circles are results for individual listeners. RT is plotted excluding outliers (see Experiment 1 analysis).

Figure 3.2C shows results for RT (excluding outliers – see Analysis section) by group and condition. Using a generalized linear mixed model on a gamma distribution for RT, we found fixed effects of Condition, *F*(4,1111) = 21.64, *p* < .001, and Melody, *F*(9,1111) = 3.80, *p* < .001, but no other effect or interaction. For Condition, pairwise comparisons revealed that pitch and stretched pitch both produced lower RTs than any of the other three conditions. For Melody, pairwise comparisons revealed that only *Pomme de reinette et pomme d'api* produced significantly lower RTs than average. Figure 3.3C shows results for RT by melody.

Discussion

On a closed-set familiar melody recognition task testing LTM of melodic contour in different dimensions, we found that amusics were significantly impaired relative to controls overall. This is in line with previous findings that difficulty in recognizing melodies without lyrics is a defining feature of amusia, despite normal ability to recognize spoken lyrics (Ayotte et al., 2002). However, across both groups, every condition produced better accuracy than rhythm alone, and there was no interaction of group by condition, suggesting that both amusics and controls are capable of using brightness, loudness and pitch cues to hear the contour of a familiar melody. The pitch condition did not produce significantly better accuracy than the stretched pitch condition in either group, though the trend is in the expected direction for both groups. This finding suggests no clear benefit of tonality cues over the contour-only conditions of stretched pitch, brightness, and loudness, unlike the result from McDermott et al. (2008). This tonality effect is possibly absent in our study because the closed-set nature of our task resulted in performance that was close to ceiling.

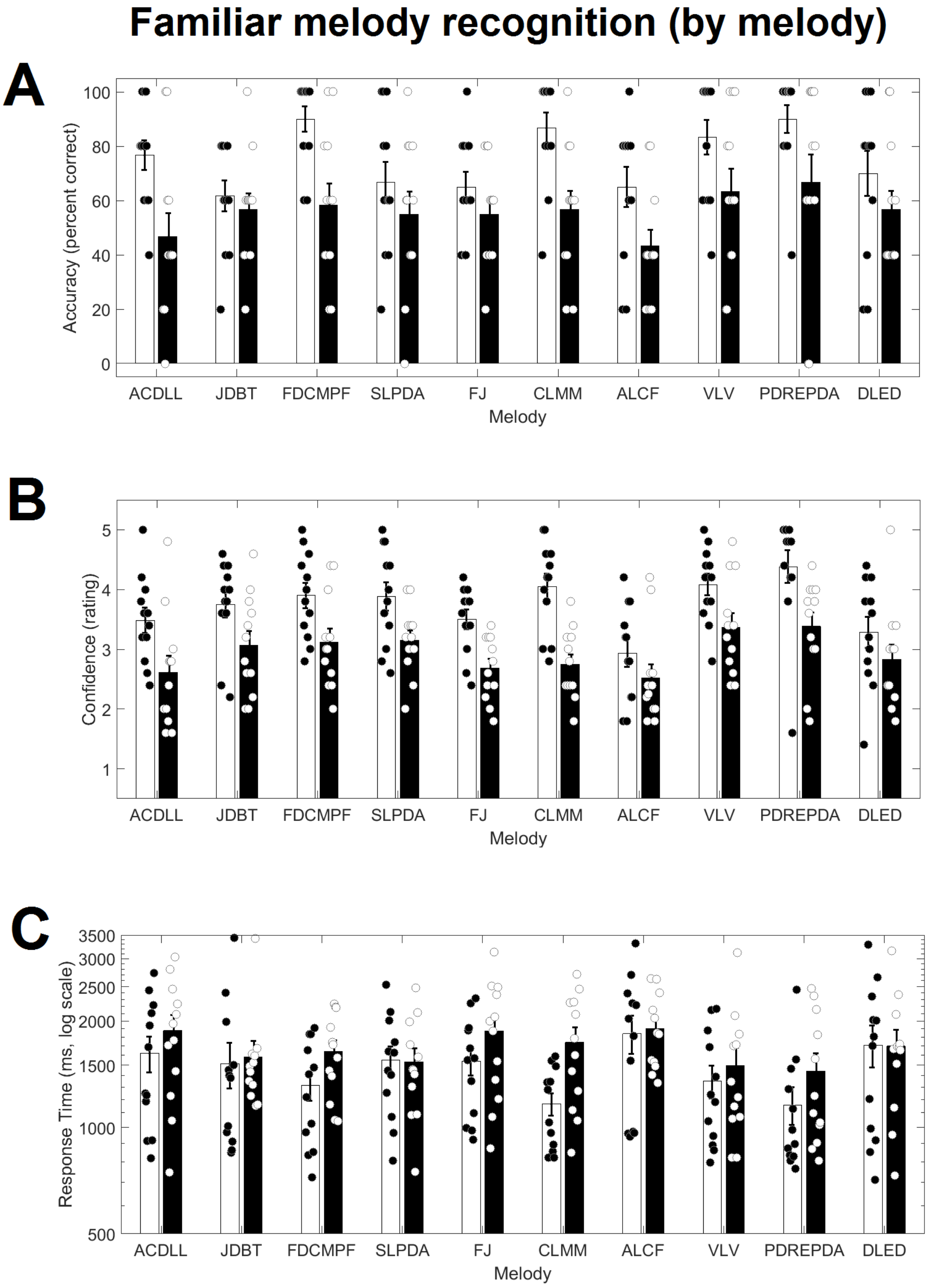


Figure 3.3. Results from Experiment 3.1, by melody, for accuracy (top), confidence (center), and RT (bottom). Legend as in Figure 3.2.

The results for subjectively rated confidence were similar to those for accuracy, but an interaction between group and condition was found, revealing a pattern where amusics were less confident than controls on the three contour-only conditions, but not significantly less confident for correct pitch or rhythm only. This finding suggests that amusics may feel particularly uncertain when presented with unnatural transformations of melodies, even though they are not more impaired at recognizing these transformations than at recognizing the more common presentations of pitch or rhythm only. Indeed, there is evidence to suggest that the mechanical, synthesized timbres often used in psychophysical studies like the present study may have a detrimental effect on memory for melodies, since these melodies were originally learned with full expressive cues (Tillmann et al., 2013). Furthermore, a drastic change in timbre has been shown to negatively affect recall of memories from LTM (Schellenberg & Habashi, 2015). It is therefore possible that performance by both groups in this experiment does not fully reflect their memory capabilities if tested with more naturalistic stimuli. Since amusics were especially unconfident in the most unnatural or uncommon conditions, it seems possible they are especially susceptible to this problem. Future research could resolve this by removing some of the facilitating aspects of the present study (training, closed set recognition) but using more ecologically valid stimuli, including the timbre and expressivity of a real-world musical performance.

For RT, no significant effect of group was observed, although effects of melody and condition were found. This suggests that amusics were as quick as controls to respond, even though they were less accurate and less confident. If we interpret RT as a measure of implicit processing (Albouy et al., 2013; Tillmann et al., 2014), the relatively fast RTs of amusics may reflect an implicit familiarity with the melodies even when they fail to accurately identify them. On the other hand, the results from subjective confidence ratings do not agree with this interpretation. Since subjects could not respond until the melody finished playing, and responses were generally slow as they involved pointing and clicking with a mouse, this RT measure may simply not be sensitive enough to detect a group-level difference. However, it was sensitive enough to detect differences between conditions and melodies.

Effects of melody were included in the models mainly to improve the fit of the models, not because there was any *a priori* reason to assume that specific melodies would produce better or worse performance. *Pomme de reinette et pomme d'api* produced the highest accuracy, and produced significantly higher confidence and lower RT than the average melody. This is possibly due to its distinctive large interval jumps, including several octave leaps. Notably, *Fais dodo, Colas, mon p'tit frère* did not stand out for good or bad performance, despite the fact that it was the only melody subdivided into three instead of two. If listeners were focusing on meter as a cue, this melody would be very easy to distinguish from the other nine. A previous study found strong effects of rhythmic segregation on familiar melody recognition in the presence of an interleaving distractor melody (Devergie et al., 2010). That study used isochronous versions of familiar melodies in order to remove rhythm cues within the melodies. An interesting future extension of the present study would involve creating isochronous contour-only melodies (i.e. isochronous melodies in stretched pitch, brightness, or loudness) to evaluate the relative contributions of rhythm and contour cues.

IV. Experiment 3.2: Short-term memory: novel contour discrimination

Method

In Experiment 3.2, as in Experiment 3.1, we sought to simplify a task from McDermott et al. (2008) in order to increase the likelihood of observing above-chance performance in our amusic group when evaluating performance using melodies defined in pitch, brightness, or loudness. The task in Experiment 3.2 involves STM rather than LTM by requiring the discrimination of novel melodic contours.

In order to reduce the load on STM, we reduced the length of the melodies from the 5 notes used by McDermott et al. (2008) to 4 notes, as used in several other studies (e.g. Lau, Mehta, & Oxenham, 2017; Oxenham et al., 2011; Pressnitzer, Patterson, & Krumbholz, 2001). Because longer durations in musical testing material tend to produce better melody recognition in general (e.g. Dowling & Tillmann, 2014), and amusics specifically may do better with stimuli of longer durations (Albouy, Cousineau, Caclin, Tillmann, & Peretz, 2016), we increased the duration of each tone to 500 ms.

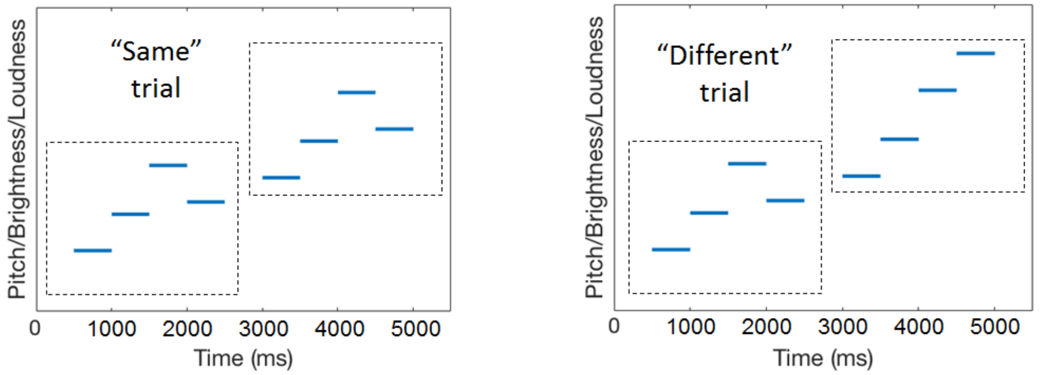


Figure 3.4. Schematic illustration of the two trial types for Experiment 3.2. Half of all trials were "same" trials, where the second melody was an exact transposition of the first. The other half were "different" trials, where one interval was inverted in the second melody. The inverted interval was either the 2nd or 3rd of the three intervals. In this example, the change occurs in the third interval (down in the first melody, up in the second melody).

Stimuli and Procedure. Each listener completed 9 blocks of 16 trials each, with 3 blocks for each condition (pitch, brightness, and loudness), in pseudorandom order such that no condition was ever repeated. This produced 48 total trials per condition per subject. On each trial, listeners heard two 4-note melodies and were asked whether their contours were the same or different, with four response options: “sure different”, “different”, “same”, and “sure same.” Figure 3.4 illustrates a “same” and “different” trial.

One step was defined as 3 dB in level, 2 semitones in spectral centroid, or 1 semitone in F0, for the loudness, brightness, and pitch conditions, respectively. Each melody contained 4 notes, thus 3 intervals between the notes. The intervals in S1 (the first melody) were randomly drawn (with replacement) from the set of [-4 -3 -2 -1 +1 +2 +3 +4] steps. The intervals in S2 (the second melody) were either identical to those in S2 or one and only one of the intervals was different.

A block contained 16 trials. On 8 of those sixteen trials, S2 had the same intervals as S1. The other 8 trials each had a different kind of change: either the 2nd or the 3rd interval out of 3 was changed, the change was either up or down, and the changed interval was either big or small. Based on pilot testing, “big” and “small” changes were determined to be 6 and 5 steps in loudness, 5 and 4 steps in brightness, or 2 and 1 steps in pitch, respectively. For balance, on all 16 trials, the 2nd or 3rd interval of S1 was also changed to be either the “big” or “small” change size, either identical to, or exactly opposite to, the corresponding interval in S2.

Finally, the intervals were converted to absolute values with a transposition between melodies, where S2 was presented in a higher range than S1. In pitch, the highest note of S1 was always 337 Hz, and the highest note of S2 was always 476 Hz. In pitch, the brightest note of S1 always had its centroid at 1588 Hz, and the brightest note of S2 always had its centroid at 3175 Hz. In loudness, the loudest note of S1 was always 62 dB SPL, and the loudest note of S2 was always 80 dB SPL.

Analysis. Each of a subject's 48 trials could be uniquely identified with a combination of 6 independent variables: block number (1-3), condition (pitch, brightness, loudness), trial type ("same" or "different" trial), size of change ("big" or "small"), position of change (2nd or 3rd interval), and direction of change (up or down). The "change" variables were still meaningful on "same" trials, since even on those trials, one interval was reassigned, simply with the same sign in the two melodies rather than opposite signs. The subject's response was coded as two logistic variables: accuracy (right or wrong), and confidence ("sure" or "not sure").

We then modeled the variance for both accuracy and confidence by fitting a GLMM in SPSS ("IBM SPSS Statistics for Windows, Version 24.0.," 2016), with the 6 fixed factors delineated above, and Subject as a random factor. We also included all potential two-way interactions between Group, Condition, Block, and Trial Type, but no other two-way or higher-level interactions. For both accuracy and confidence, the response distribution was treated as binomial, with a logit link function.

Results

Figure 3.5A shows results for accuracy, by group and condition. Using a generalized linear mixed model on a binomial distribution for accuracy, we found fixed effects of Group, $F(1,3433) = 9.36$, $p < .001$, Condition, $F(2,3433) = 4.84$, $p = .002$, Trial Type, $F(1,3433) = 70.76$, $p = .008$, and Change Direction, $F(1,3433) = 47.32$, $p < .001$. We also found an interaction of Group by Condition, $F(2,3433) = 13.11$, $p < .001$, and

Condition by Trial Type, *F*(2,3433) = 42.29, *p* < .001. No posthoc pairwise comparisons were significant, but the difference between amusics and controls seemed smaller for loudness than for the other two conditions, and that listeners did better on “different” trials than “same” trials only for the brightness condition. Performance in all conditions for both groups was significantly above chance, based on one-sample t-tests comparing to 50% correct (*p* < .003 in all cases).

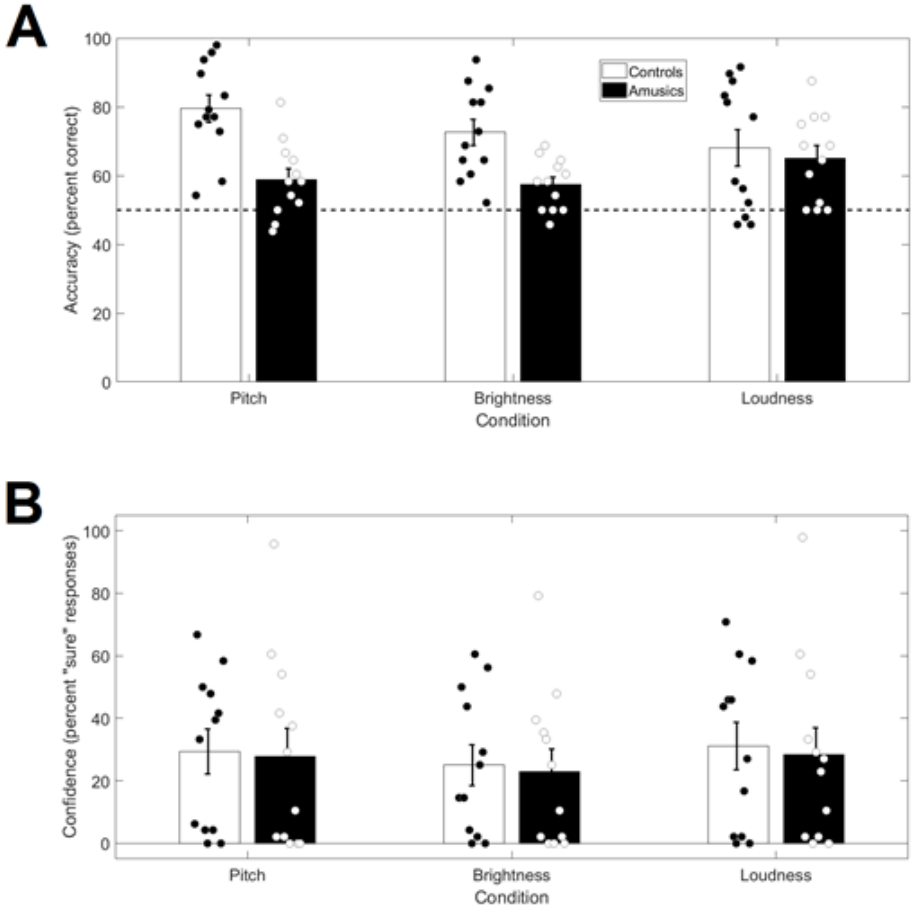


Figure 3.5. Contour discrimination results with novel melodies from Experiment 3.2, by group and condition. A. Accuracy was determined by whether or not response matched trial type. B. Confidence was determined independently of accuracy, by whether a “sure” response was selected. Error bars represent ±1 standard error of the mean. Circles are results for individual listeners.

Figure 3.5B shows results for confidence, by group and condition. Using a generalized linear mixed model on a binomial distribution for confidence, we found fixed effects of Condition, $F(2,3433) = 7.83$, $p < .001$, and Change Direction, $F(1,3433) = 17.56$, $p < .001$. We also found an interaction of Group by Block Number, $F(2,3433) = 5.17$, $p = .006$. Once again, no posthoc pairwise comparisons were significant, but on block 1, amusics seemed on average more confident than controls, but by block 3 this pattern had reversed. This effect can be observed in Figure 3.6B, showing results for confidence by block. (The Group by Block Number interaction was not significant for accuracy; Figure 3.6A shows these results for comparison with Figure 3.6B).

Discussion

The results from the control group replicate the results of McDermott et al. (2008): novel melody contour discrimination is possible in all three dimensions of pitch, brightness, and loudness. More interestingly, although amusics were impaired relative to the controls, their mean performance was significantly above chance in all three conditions. In addition, the difference between amusics and controls was observed not just for pitch but also for brightness, although the results in loudness seemed more closely matched between the two groups. Interestingly, while amusics are impaired relative to controls, they may not be any less confident, since we observed no fixed effect of group on confidence. Instead, as revealed by the interaction between group and block number for confidence, amusics became less confident over the course of Experiment 3.2, while controls grew more confident.

Change direction had a strong effect on both accuracy and confidence, with subjects doing better for upward changes than downward changes. This may be simply a function of upward leaps being more salient and unusual than downward leaps, especially at the end of a melody. Subjects were also generally biased towards caution, responding "same" more often than "different", as evidenced by the effect of trial type, but this pattern is reversed in brightness.

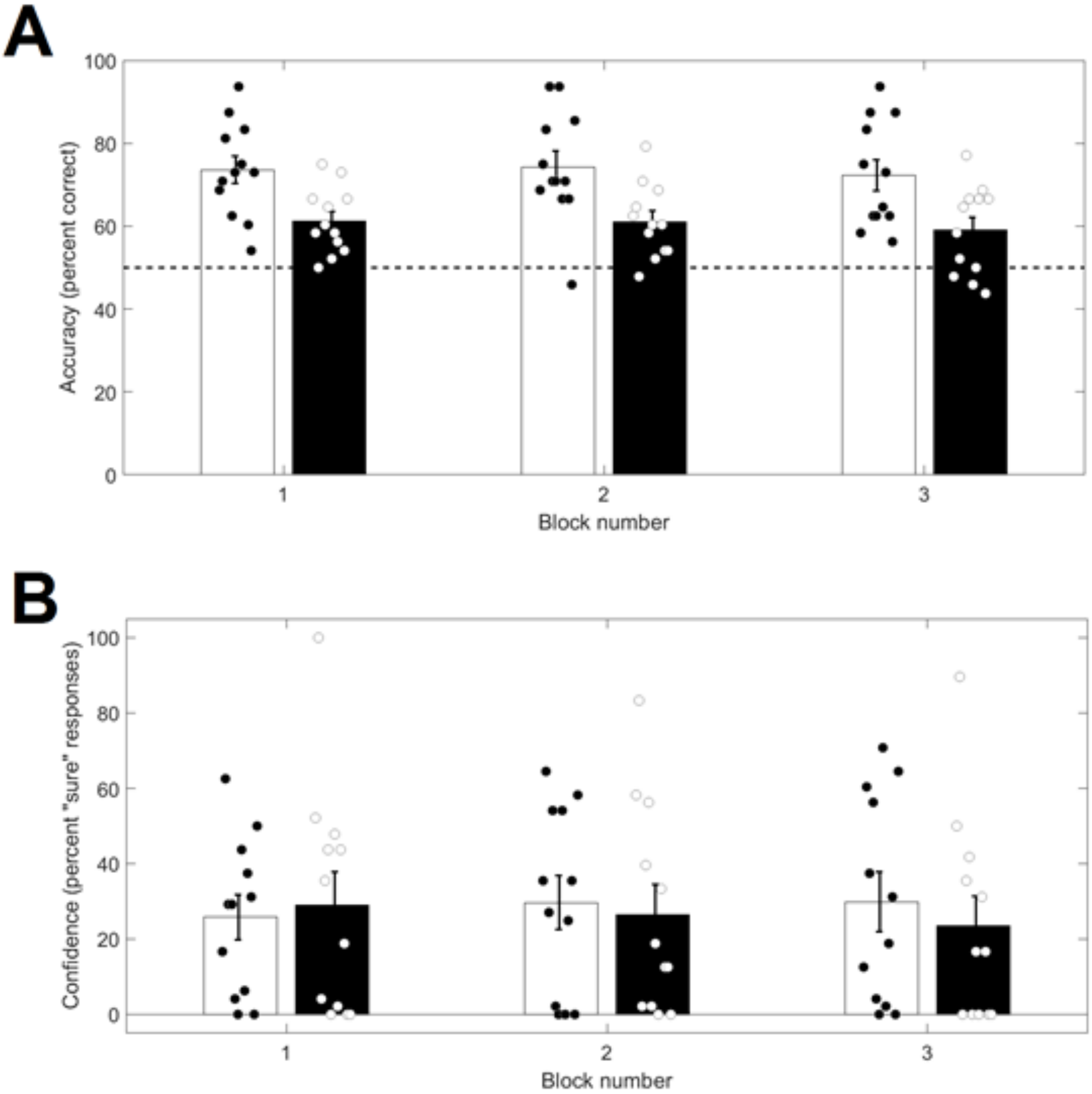


Figure 3.6. Results from Experiment 3.2, by block number (in terms of the order the blocks were completed). Each block contained three runs, one for each auditory dimension (pitch, brightness, and loudness). The order of the auditory dimensions was pseudorandom on each block, with the constraint that no dimension was repeated over the course of the experiment. Here, results are pooled across the three auditory dimensions used. Legend as in Figure 3.5.

The most interesting result in Experiment 3.2 is the finding that amusics are not equally impaired for pitch, brightness, and loudness contours, as shown by the interaction of group by condition on accuracy (see Figure 3.5A). It appears that their impairment on loudness contour discrimination specifically is less pronounced than for pitch and brightness. This agrees with the recent finding that amusics are less impaired at cross-modal integration with brightness or loudness than with pitch (Lu et al., 2017), although in this case, their lack of impairment comes within a single dimension, rather than in the context of audiovisual integration. Because loudness is generally unrelated to pitch, this finding offers some support for the theory that the deficits in amusic are specific to at least pitch-related dimensions (including brightness). This conclusion seems a little less sweeping than the conclusion from Experiment 3.1, which suggested that amusic performance was equally affected in all three dimensions tested. This apparent discrepancy is addressed in the final section of this chapter.

V. General discussion

We measured performance of amusics and controls on two contour memory tasks, one involving LTM (Exp. 3.1) and one involving STM (Exp. 3.2). The tasks involved the auditory dimensions of pitch, brightness, and loudness. The results from the control participants were consistent with the results from McDermott et al. (2008). For familiar melodies, recognition was best with the correct pitch relationships and was poorer, although still above chance, when the melodies were defined in terms of stretched-pitch, brightness, or loudness contours. Poorest performance (but still above chance) was observed when only rhythm cues were provided. For unfamiliar melodies (testing STM), a similar pattern of results was observed, with performance possible for all three dimensions tested but with progressively worse performance for pitch, brightness, and loudness conditions.

The results from the amusic group for familiar melodies showed the same pattern of results as control subjects, though with impairment in every condition. For unfamiliar melodies, however, the pattern of results was different from that of both control subjects in this study and previous results from McDermott et al. (2008). The best performance was observed for loudness, not pitch.

Overall, the results suggest that amusics are capable of extracting and storing melodic contours in both STM and LTM, even if their ability to do so is impaired relative to controls, and that they share the ability of control participants to extract contour from brightness and loudness patterns.

There is an apparent discrepancy between the results from Exp 3.1, where amusics showed roughly equivalent impairment in all conditions, and from Exp 3.2, where impairment was greater for pitch and brightness than for loudness. This may be explained by the fact that in Exp. 3.1, the stretched-pitch, brightness, and loudness conditions all depended on a LTM representation based on the original pitch of the familiar melodies. If this original pitch representation is degraded in amusics, then it makes sense that it would be degraded when tested in all other dimensions, regardless of whether those other dimensions are affected by amusia. However, in Exp. 3.2, the brightness and loudness conditions require no processing or memory of pitch at all, and so could in principle be unaffected by any pitch-selective deficits.

A potentially promising future direction is to more directly test importance of tonality in LTM and STM for amusics and controls, since there was inconclusive evidence for tonality perception in the present study. Using the same tasks as the present paper (familiar melody recognition and/or novel contour discrimination), one could create stimuli that contain “wrong notes” which violate tonal structure, contour, both, or neither, in the style of Dowling (1978). It would be informative to see what kinds of errors impair recognition most for amusics, and whether it differs for LTM and STM-based paradigms.

Chapter 4: Melodic and harmonic tonal context in pitch interval perception

Sections I-IV are reprinted from:
Graves, J. E., & Oxenham, A. J. (2017). Familiar tonal context improves accuracy of pitch interval perception. *Frontiers in Psychology, 8*(October), 1753.

I. Introduction to pitch interval perception

Pitch, a primary dimension of auditory sensation, is an attribute closely related to the fundamental frequency (F0) or overall periodicity of a sound. In speech, rising and falling pitch contours serve as cues to a speaker's emotions, intentions, and emphasis, and as cues to semantic meaning in tonal languages. In music, sequences of pitch define melody and simultaneous combinations of pitch define the harmony of chords. In Western music, as in many other traditions, pitches are organized into discrete categories within a tonal hierarchy such as a musical key. Listeners, especially those with musical training, are sensitive to these hierarchies, rating some notes as better "completions" than others following a musical scale (Krumhansl & Shepard, 1979), or following a single chord or sequence of chords (Krumhansl & Kessler, 1982). The resulting "tone profiles" of perceived pitch relationships within the key cannot be predicted simply from proximal stimulus similarities, and instead are thought to reflect prior knowledge and exposure (Parncutt & Bregman, 2000). Tonal structure is a strong factor influencing psychological expectancies for both melody and harmony (Schmuckler, 1989). For melodies, listener expectations are also heavily influenced by contour (Cuddy & Lunney, 1995), in accordance with contour-based models (Narmour, 1990). Thus, to fully describe listener expectations for melodic continuation, it is necessary to consider both tonal structure and melodic contour as separate influences (Graves et al., 2014).

Sensitivity to tonal hierarchies may be the result of a process of statistical learning, wherein listeners come to expect musical patterns to which they have frequently been exposed. Statistical learning for pitch patterns has been observed on a small scale in both infants and adults (Saffran, Johnson, Aslin, & Newport, 1999), in a process analogous to learning of word segmentation in language development. On a larger scale, tonal expectations in Western listeners are well explained by statistical regularities in familiar Western music such as folk songs and chorales (Pearce & Wiggins, 2006). This learning likely happens very early in life, as infants as young as 7 months are sensitive to familiar tonal structures (Cohen, Thorpe, & Trehub, 1987). However, specialization for a particular tonal hierarchy may take time to develop fully: while 6-8 month-old infants are equally able to detect violations of various tonal structures, Western adults are especially sensitive to violations of the Western diatonic scale (Lynch, Eilers, Oller, & Urbano, 1990; L J Trainor & Trehub, 1992).

Once learned, tonal sensitivity is a robust phenomenon. Familiar melodies are stored in long-term memory based on tonal structure, not only contour (Dowling & Fujitani, 1971), and even short-term memory for novel melodies is influenced by tonality (Boltz, 1991; Dowling, 1978). In fact, for musically trained listeners, tonal hierarchies need not even be cued physically: tone profiles of pitch relationships within a musical key can also be measured following imagined (not physically presented) tonal hierarchies (Vuvan & Schmuckler, 2011).

Accessing these overlearned tonal hierarchies can facilitate pitch processing when the relevant pitches are highly expected within the tonal structure. For various pitch processing tasks, response times are faster for expected than for unexpected chords, based on the preceding harmonic progressions (Bharucha, 1987; Bigand & Pineau, 1997; Tillmann, Janata, Birk, & Bharucha, 2008; Tillmann & Marmel, 2013), as well as for notes primed by melodic context (Marmel, Perrin, & Tillmann, 2011; Marmel, Tillmann, & Dowling, 2008). The mechanism of this facilitation of processing may be either priming of expected pitches, or enhanced perception and representation of important pitches or harmonies within the tonal hierarchy. Under the former expectation-based explanation, a pitch that is predicted or expected by a tonal hierarchy may produce a faster response time simply because less time is required to react to an unsurprising or predictable event. This explanation is favored by most reaction-time studies, e.g. Bigand & Pineau (1997). Under the second perceptual-accuracy-based explanation, however, response times could be faster with tonal context because the representation of pitch at some level in the auditory system becomes more accurate, rendering the task easier. This may take the form of anticipatory activation of expected pitches (e.g. Bharucha, 1987). Increased event-related potential (ERP) amplitudes to pitches high in the tonal hierarchy provide additional evidence for enhanced neural representation of these pitches (e.g. Krohn, Brattico, Välimäki, & Tervaniemi, 2007).

If the decreased response times in these studies reflect an enhanced sensory representation of pitch, we might expect to observe improvements in measures of performance or accuracy as well. One such measure is pitch discrimination, where the listener directly compares two pitches presented in sequence. For this task, tonal context has been found to improve accuracy, but the observed effects have been small relative to effects on response time, and in some cases may be modulated by differences in timbre between tones (E. Borchert & Oxenham, 2010; Marmel et al., 2008; Warrier & Zatorre, 2002). Harmonic priming studies have used a dissonance detection task in which the listener must detect the presence of an augmented root or augmented fifth (both highly dissonant chord members in Western music). Tonal context also has small and inconsistent effects on accuracy for this task (Bigand & Pineau, 1997; Tillmann & Marmel, 2013), in contrast with robust effects on response time. However, in these tasks, the mistuning can be detected also by the presence of acoustics beats in the waveform of the dissonant interval (McDermott, Lehr, et al., 2010), meaning that the pitch interval itself need not be discriminated by the participants. Thus, the lack of a robust effect of tonal context on task accuracy in these situations does not necessarily imply a lack of pitch enhancement through tonal context.

It is possible that tonal context effects are stronger for pitch *interval* discrimination than for simple pitch discrimination. Pitch intervals determine tonal hierarchies and set pitch apart from other auditory dimensions such as brightness and loudness (Graves et al., 2014; McDermott et al., 2008). Interval discrimination may be a more difficult task, due to the higher cognitive load required to represent distances (intervals) as opposed to individual values (pitches) in working memory. This could be the reason that discrimination thresholds, or difference limens (DLs), for pitch intervals are large compared to basic pitch DLs, which are exceptionally low among auditory dimensions (McDermott, Keebler, et al., 2010). With more room for improvement, one might expect that any enhancement of the sensory representation of pitch would be especially beneficial on a pitch interval perception task. In addition, one known effect of tonal structure on pitch interval perception is that tonality allows for categorical perception of discrete intervals, as opposed to a continuous range. There is some evidence that musicians may more accurately discriminate pitch intervals at category boundaries than within an interval category, though this effect is not robust, and is sensitive to differences in experimental methodology (Burns & Ward, 1978). However, the effect was not observed at all in non-musicians, suggesting that categorical perception, if present, is learned. In a convergent finding, small frequency oscillations are more easily detected when centered around perfect octaves and fifths than neighboring intervals (L Demany & Semal, 1992). The subjective "octave" category is slightly stretched relative to a physical octave (doubling in frequency), but approaches the physical octave when tonal context is introduced (Cuddy & Dobbins, 1988). Activating a tonal hierarchy could potentially enhance pitch interval perception by sharpening distinctions between primed interval categories. In other words, within a tonal context, a musical interval that is larger or smaller than expected may result in the second note being perceived as a "sour note" with respect to its expected pitch value, rather than in terms of the interval size between it and the preceding note.

A previous study found that the discrimination of musical intervals was better following a short melody than for intervals presented in isolation, suggesting that tonal context does enhance perception of pitch intervals (Wapnick, Bourassa, & Sampson, 1982). However, certain aspects of that study's methodology leave its results open to interpretation. Firstly, only participants with a very high degree of musical experience were tested, and these participants received additional extensive training on an interval labeling task before completing the interval discrimination task. Many of them reported having absolute pitch, and all of them showed some degree of absolute pitch labeling ability, making it unclear whether the participants even needed to compare the two tones in each trial to complete the task. Although benefit from melodic context should not depend on absolute pitch possession, this may have transformed the putative relative-pitch task into functionally an absolute-pitch task. Secondly, the first pitch of the first interval on discrimination trials was always held constant, potentially allowing participants to use absolute pitch, instead of relative pitch, and so employ basic pitch discrimination instead of pitch interval discrimination. Thirdly, no distinction was made between a musical context that defines a congruent tonal hierarchy (such as a major key) and a tonally incongruent musical context: participants heard either a familiar melody or nothing. Thus, the benefit of a tonal context may be due to the reinforcement of tonality, or simply due to the presence of any context pitches, regardless of their tonal congruence.

The present study sought to determine whether a prior tonal context enhances pitch representations in a way that improves pitch interval discrimination. In order to ensure that participants were perceiving relative pitch intervals, we roved all absolute pitches in the study across a continuous range of fundamental frequencies. To dissociate various potential interpretations of a difference between familiar melodic context and no context, we also included three control conditions: a Repetition condition to test the effect of simply reinforcing the target pitch without any reference to a tonal (e.g., major or minor) center, and two unfamiliar melodic contexts (Mistuned and Whole-Tone Scales) for comparison with the more familiar (diatonic, Major Scale) context. If familiar tonal hierarchies do in fact facilitate pitch processing by enhancing the sensory representation of pitch or pitch intervals, we would predict that tonal context improves performance in an interval discrimination task, but only in cases in which the context provides congruent tonal cues. The first experiment established context using a melodic sequence of single pitches. The second experiment established context using a harmonic sequence of multiple pitches in the form of an authentic cadence. In all cases, care was taken to ensure that none of the context tones was of the same pitch class as the test tone itself, to avoid the possibility that participants were making a direct comparison between the test tone and one of the context tones.

II. Experiment 4.1: Melodic context for pitch intervals

Method

Stimuli. Participants heard sequences of pitches carried by harmonic complex tones. The tones were generated with all harmonics lower than the Nyquist frequency (22.05 kHz), and were lowpass filtered with a cutoff frequency of 200 Hz and a -12 dB/octave slope. The overall level of each tone after filtering was 60 dB SPL. The tones were generated within MATLAB (The Mathworks, Natick, MA), using a 24-bit L22 soundcard (Lynx Studio Technology, Costa Mesa, CA), presented diotically through HD650 headphones (Sennheiser USA, Old Lyme, CT) at a sampling rate of 44.1 kHz.

Figure 4.1 shows the paradigm for stimulus presentation in the five melodic context conditions. The task-relevant stimuli on each trial were two tones presented sequentially, each with a duration of 400 ms, including 10-ms raised-cosine rise and fall ramps, separated by a gap of 100 ms. Trials in the No Context condition consisted only of the test interval formed by these two test tones. The F0 of the first test tone was randomly chosen from a uniform distribution within a 1.5-octave range from 200 to 565.69 Hz (approximately G3 to C#5). On half of the trials, the second test tone's F0 was higher than that of the first by a ratio exactly equal to a standard interval in the diatonic equal-tempered scale. On the other half of trials, the second test tone's F0 was higher than the frequency that would be chosen by the standard interval size; we termed the ratio of this discrepancy $\Delta F0$. Thus, if the test tone was 1 semitone higher than the frequency that would have been selected in the standard interval, ΔF0 would be approximately 6% ($2^{1/12}$). Participants were instructed to judge whether each interval was "small, in tune" (when the F0 difference was exactly the standard interval size) or "large, mistuned" (when the F0 difference was greater than the standard interval size by $\Delta F0$). In this way, participants had the option of using either the size cue ("small" or "large") or the tuning cue ("in tune" or "mistuned") to complete this task. Two standard interval sizes were tested in two separate phases of the experiment. These were 2 semitones (a major second) or 5 semitones (a perfect fourth) in the equal-temperament tuning system. We chose common intervals in Western tonal music because our participants were more likely to have been exposed to Western tonal music than other musical styles. By avoiding standard intervals larger than 5 semitones, we avoided repeating any pitch classes from the context sequence.

On trials in the other four conditions, the test interval was preceded by a melodic context sequence, consisting of four tones with durations of 400 ms each (including 10-ms raised-cosine rise and fall ramps), separated by 100-ms gaps, with 600 ms silence between the context sequence and the final two test tones. In each context condition, the F0 of the final tone in the context sequence was equal to the F0 of the first tone in the test interval. In the Repetition condition, all four context tones had the same F0 as the first test tone. In the Mistuned condition, each tone in the context sequence was exactly 1.5 semitones higher than the previous tone. In the Whole Tone condition, each tone in the context sequence was exactly two semitones higher than the previous tone. Finally, in the Major condition, the context sequence corresponded to a major (diatonic) scale ascending from the dominant (5th) scale degree to the tonic, with successive interval sizes of two semitones, two semitones, and one semitone. These four context conditions were designed to dissociate the effects of pitch reinforcement, directional context, tuning cues, and tonal hierarchy, respectively. All four context conditions provide some pitch reinforcement: additional examples of the first pitch of the test interval may be helpful. Mistuned, Whole Tone, and Major conditions all provide directional context: upward intervals of fixed size are presented, against which the test interval could be compared. Only Whole Tone and Major conditions fit within the Western 12-tone chromatic scale, and only the Major condition fits within the Western hierarchical diatonic scale.

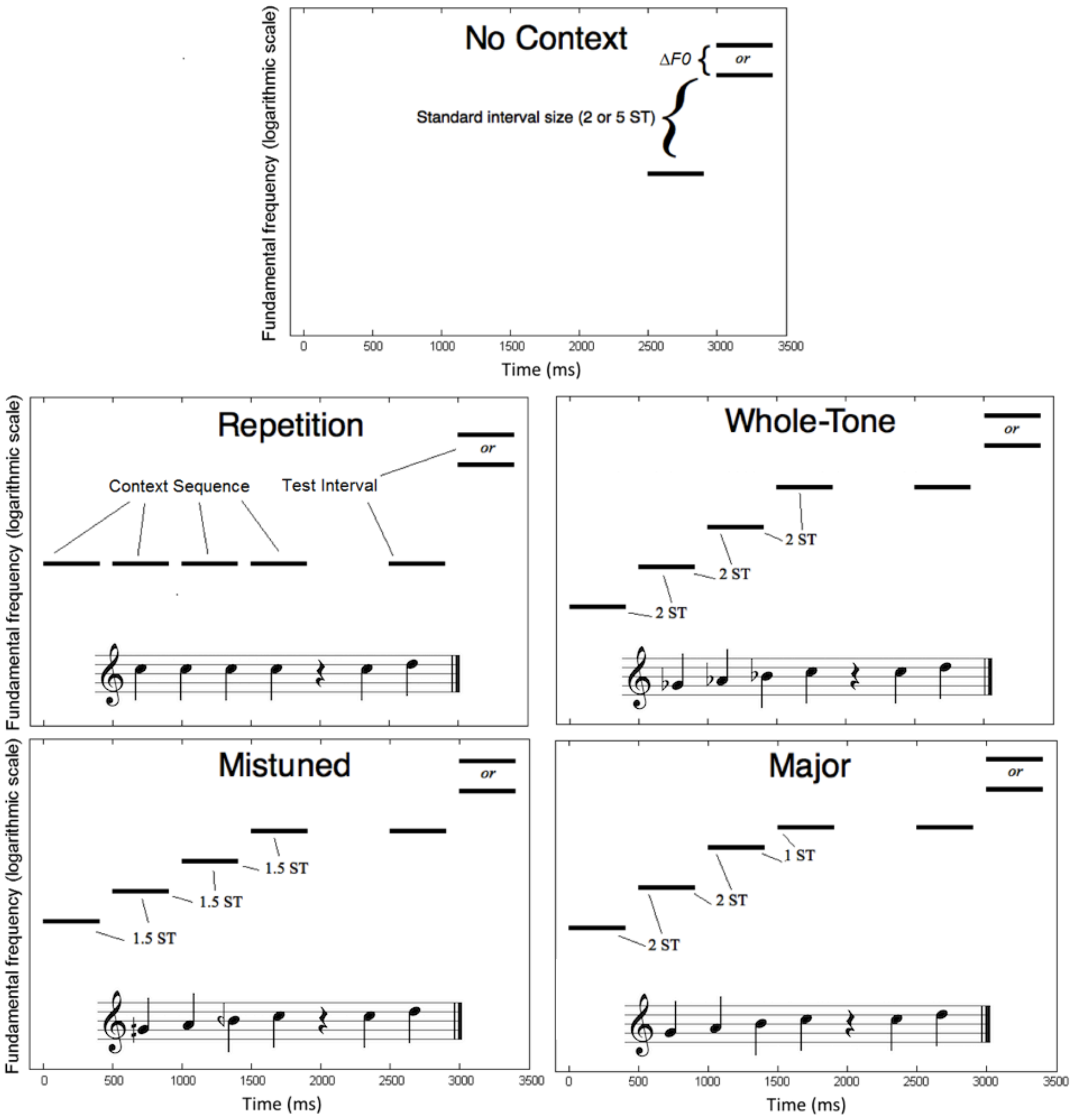


Figure 4.1. Schematic diagrams in spectrogram form of a single trial in each of the five conditions for Experiment 4.1 (melodic context). Pitch distances are labeled in semitones (ST). Context conditions are also illustrated with musical notation.

On trials in the other four conditions, the test interval was preceded by a melodic context sequence, consisting of four tones with durations of 400 ms each (including 10-ms raised-cosine rise and fall ramps), separated by 100-ms gaps, with 600 ms silence between the context sequence and the final two test tones. In each context condition, the

F0 of the final tone in the context sequence was equal to the F0 of the first tone in the test interval. In the Repetition condition, all four context tones had the same F0 as the first test tone. In the Mistuned condition, each tone in the context sequence was exactly 1.5 semitones higher than the previous tone. In the Whole Tone condition, each tone in the context sequence was exactly two semitones higher than the previous tone. Finally, in the Major condition, the context sequence corresponded to a major (diatonic) scale ascending from the dominant (5th) scale degree to the tonic, with successive interval sizes of two semitones, two semitones, and one semitone. These four context conditions were designed to dissociate the effects of pitch reinforcement, directional context, tuning cues, and tonal hierarchy, respectively. All four context conditions provide some pitch reinforcement: additional examples of the first pitch of the test interval may be helpful. Mistuned, Whole Tone, and Major conditions all provide directional context: upward intervals of fixed size are presented, against which the test interval could be compared. Only Whole Tone and Major conditions fit within the Western 12-tone chromatic scale, and only the Major condition fits within the Western hierarchical diatonic scale.

Participants. Twenty-one participants, 9 male and 12 female, were recruited from the Twin Cities campus of the University of Minnesota. They ranged from 18 to 25 years of age (*Mean* = 19.8, *SD* = 1.9), and from 0 to 15 reported years of musical experience (*Mean* = 5.9, *SD* = 5.2), with musical experience loosely defined as regularly playing any musical instrument. All participants were screened for normal audiometric hearing thresholds, defined as not exceeding 20 dB hearing level (HL) for frequencies between 250 and 8000 Hz. This study was approved by the University of Minnesota Institutional Review Board, as part of the study titled 'Complex Pitch Perception in Complex

Environments', protocol number #0605S85872. The experiment was completed in a single 2-hour session per participant. All participants provided written informed consent and were compensated $10 per hour for their participation.

Procedure. To allow the participants to gain familiarity with the standard interval size, each participant completed the entire experiment for one standard interval before being tested on the other standard interval. The order of the standard interval presentation was counterbalanced between participants, such that 11 participants completed the procedure for the 2-semitone standard first, while 10 participants completed the procedure for the 5-semitone standard first.

Because the task was novel and not intuitive for many participants, each participant began with orientation and training before moving on to the testing phase. The orientation phase consisted of listening to 5 labeled examples of the small, in-tune interval and 5 examples of the large, mistuned interval. For this demonstration, the $\Delta F0$ ratio was fixed at 8% (larger than a semitone). During the orientation phase, participants did not respond, but merely listened to the labeled examples.

The training phase consisted of 3 blocks of 40 trials each in the No Context condition. For the first block, $\Delta F0$ was fixed at 12.6% (just larger than 2 semitones). For the second block, $\Delta F0$ was 8%, and for the third it was 5% (just under a semitone). No time limit was imposed on responses during this training period. Participants generally performed near ceiling during this training phase, making few errors, as the $\Delta F0$s used were large.

Following training, each participant's DL for $\Delta F0$ in the No Context condition was estimated in a pilot phase of the experiment using an adaptive tracking procedure.

The geometric mean estimated DL was 3.1% for the 2-semitone standard, 95% CI [2.3% 4.2%], and 2.8% for the 5-semitone standard, 95% CI [2.0% 4.0%]. This wide range of thresholds is typical for frequency discrimination tasks, as recently illustrated in a study of 100 participants with normal hearing (Whiteford & Oxenham, 2015).

The estimated DLs, determined for each participant individually, were used to set the Δ*F0* in the main testing phase of the experiment. Based on pilot testing, participants were expected to perform at sensitive levels (below ceiling and above chance) when tested with Δ*F0* set to roughly ¼ the threshold estimated by the adaptive tracking procedure. This discrepancy may be due to learning occurring over the course of the experiment. Accordingly, each participant was tested with Δ*F0* set to 25% of his or her initially estimated threshold. Thus, Δ*F0* was constant for each participant for each standard interval size, but different across participants and standard interval sizes according to the estimated DL.

The testing phase for each standard interval condition consisted of 25 blocks of 20 trials each. Each block contained trials with one of the five context conditions. On each trial, participants were presented with the stimulus and asked “Which kind of interval – small, in-tune or large, mistuned?” Participants were required to indicate their response via key press within 1 second of stimulus offset. The time limit was introduced in order to prevent mental rehearsal of the stimulus following the presentation. If a participant failed to respond within this time limit, the experiment program recorded a response of “small, in-tune” and proceeded to the next trial. Since this was the correct response on half of the trials, running out of time gave the participant a 50% chance of being correct. Participants were instructed to avoid running out of time, and accordingly this happened rarely: the percentage of trials on which a participant ran out of time ranged from 0.2% to 6.4% (*Mean* = 1.79%, *SD* = 1.50%).

Each participant completed five blocks for each of the five context conditions during this phase. The context condition varied from block to block, with the order of context conditions determined randomly for each consecutive set of five blocks. Participants were instructed to focus only on the final two tones (the test interval) if a context sequence was present. After the testing phase was completed for one standard interval condition, the procedure was repeated in its entirety for the other standard interval condition, starting with new orientation, training, and DL estimation periods.

Analysis. Individual participants' sensitivity (*d'*) was estimated by subtracting the z-scored (the inverse cumulative normal distribution function) false alarm rate from the z-scored hit rate. In this calculation, a hit was defined as correctly detecting the large, mistuned interval, while a false alarm was defined as incorrectly responding "large, mistuned" when the small, in-tune interval was presented. Figure 4.2 shows the pattern of performance across conditions for standard interval size.

The *d'* values, averaged across all participants, in the No Context condition were between 0.5 and 1, indicating that our estimates from the pilot phase successfully produced performance that was well above chance (*d'* = 0) but below ceiling (*d'* > ~2.5). A paired-samples t-test comparing *d'* values in the No Context condition for the two standard interval sizes was not significant (*p* = 0.52), suggesting that our pilot estimates of DLs in the baseline condition had been successful at targeting roughly equal levels of performance between the two standard interval sizes. Beyond that, since participants were tested at different $\Delta F0$ levels according to their individual estimated DLs, comparisons of absolute *d′* values between participants are uninformative. We analyzed the effect of all five context conditions with a repeated-measures ANOVA on these *d′* values, and ran post-hoc pairwise comparisons to determine the advantage of each condition over the baseline No Context, as well as benefit of one condition over another.

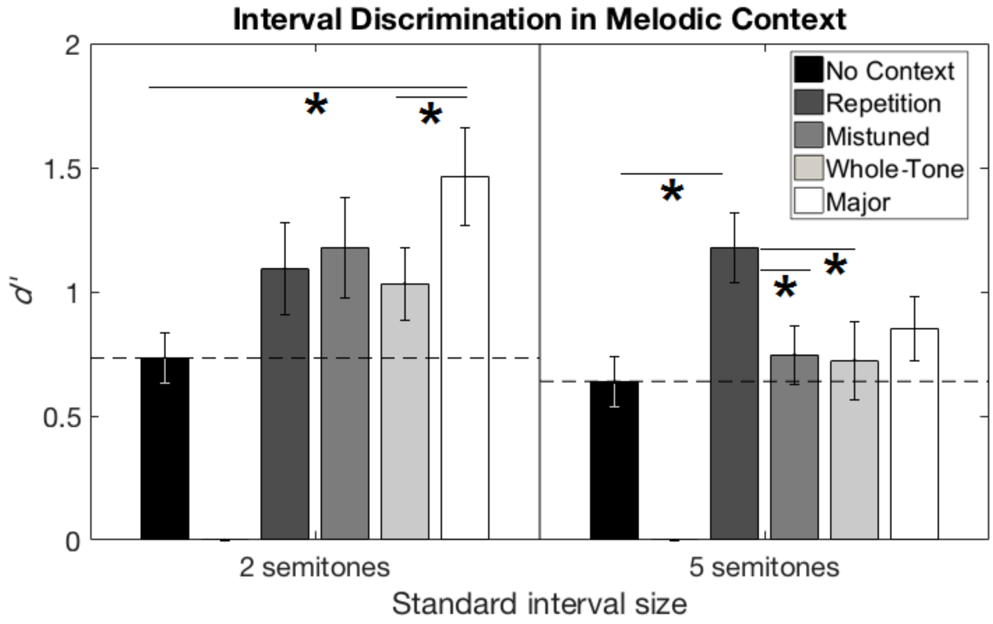


Figure 4.2. Interval discrimination performance from Experiment 1 (melodic context). Performance in *d′* is shown for the 2-semitone (left) and 5-semitone (right) standard interval sizes. Performance in the No Context condition was treated as a baseline (horizontal dashed line). Error bars represent +/- 1 standard error of the mean across participants. Horizontal solid lines with asterisks show significant pairwise comparisons between conditions for each standard interval size.

Results

The repeated-measures ANOVA on *d′* values, with two within-subjects factors of standard interval size and context condition, revealed a main effect of context condition, $F(4,80) = 10.26$, $p < .001$, $\eta^2 = .339$. Post-hoc pairwise comparisons of context conditions with Bonferroni correction (criterion $p = 0.05/10 = .005$) showed significant benefit over

No Context for Major Scale (mean difference = 0.47, $p < .001$, Cohen's $d = 1.02$) (Cohen, 1988) and Repetition (mean difference = .45, $p < .001$, Cohen's $d = .96$) contexts, as well as an advantage of Major Scale over Whole-Tone Scale context (mean difference = .28, $p = .001$, Cohen's $d = .53$). No other pairwise comparisons reached significance.

We also observed a significant interaction between context and standard interval size, $F(4,80) = 6.049$, $p < .001$, $\eta^2 = .232$. The interaction reflects in part the difference in the benefit from the Major Scale and Repetition contexts for the 2- and 5-semitone standard interval sizes. We performed 25 post-hoc pairwise comparisons to investigate this interaction: 10 comparisons between context conditions for each of the 2 standard interval sizes, and 1 comparison between standard interval sizes for each of the 5 conditions. With Bonferroni correction (criterion $p = .05/25 = .002$), for the 2-semitone standard interval size, *d'* values were higher in the Major Scale context than No Context (mean difference = .73, $p < .001$, Cohen's $d = 1.09$) or Whole Tone (mean difference = .43, $p = .001$, Cohen's $d = .90$) conditions. For the 5-semitone standard, *d'* values were higher in the Repetition context than No Context (mean difference = .54, $p < .001$, Cohen's $d = 1.04$), Mistuned (mean difference = .43, $p < .001$, Cohen's $d = 1.28$), or Whole Tone (mean difference = .45, $p = .001$, Cohen's $d = .82$) conditions. No other pairwise comparisons between conditions, nor comparisons between standard intervals within conditions, reached significance. We observed no main effect of standard interval size.

Discussion

The results of Experiment 4.1 suggest that performance on an interval discrimination task is significantly affected by the tonal context in which the task is performed. The Major Scale melodic context provided an advantage over the No Context or Whole-Tone Scale conditions, but no advantage over the Repetition or Mistuned-Scale conditions. Thus, no clear evidence was obtained for the benefit of establishing an over-learned (major-scale) tonal context over a simple repetition of the reference tone.

The interaction effect between context condition and standard interval size suggests that the pattern of improvement from context was different for the 2- and 5-semitone standard tasks. One evident difference between these patterns of results is the effect of the Repetition context and the Major Scale context in the two tasks. The best performance in the 2-semitone-standard task was from the Major Scale context, whereas the best performance in the 5-semitone standard task was from the Repetition context. In interpreting this difference, it is worth considering possible unintended tonal implications of the melodic context sequences. The intended interpretation of the Major Scale context was as the final four notes of an ascending major scale, ending on the tonic. Under this interpretation, both the 2-semitone interval and the 5-semitone interval fit in the established key. However, participants may have interpreted this sequence instead as the first four notes of an ascending major scale, beginning on the tonic. Under this interpretation, only the 2-semitone interval fits in the established key. This ambiguity may explain the reduced improvement of this context sequence in the 5-semitone-standard task.

The Repetition context, though intended as one level in a series of control conditions (disambiguating the effect of reiterating a reference pitch), could be interpreted as a repeating 5th scale degree (the dominant), anticipating the arrival of the tonic, which is exactly 5 semitones higher. This is a common pattern in traditional Western music, and the effect may have been enhanced by the rhythmic pattern established by the temporal paradigm of this experiment, such that the final tone of the test interval can be heard to fall on a downbeat. This interpretation may explain the heightened improvement of the Repetition context sequence in the 5-semitone-standard task.

If the Repetition condition had only the simple effect we intended, to reinforce the reference pitch, the simplest interpretation would be that the familiar tonal context provided an advantage over unfamiliar tonal context, but it did not provide an advantage over simple repetition of the first pitch in the test interval. This would suggest that the benefit of melodic context observed by Wapnick et al. (1982) can be disrupted with unfamiliar tonality, but may have more to do with repetition and reinforcement of target pitches than with the establishment of tonal structure. However, if we do interpret the Repetition context as inducing an accidental "tonal context" itself, these results are reasonably consistent with Wapnick et al. (1982).

III. Experiment 4.2: Harmonic context for pitch intervals

Rationale

The results of Experiment 4.1 were mixed: familiar diatonic tonal context improved performance on pitch interval discrimination over no context and one unfamiliar context, Whole Tone, but not over the other unfamiliar context, Mistuned, nor over simple tone repetition. Specifically for the 5-semitone standard interval, familiar diatonic context provided no significant advantage over no context. One possible explanation of the small degree of benefit over no context, and the lack of benefit of the familiar tonal context with the 5-semitone standard interval, is that the context of a sequence of four single tones did not establish a sufficiently strong and unambiguous sense of tonality. Indeed, many past studies have used chord progressions, rather than individual notes, to establish a clear tonality (Bharucha, 1987; Bigand & Pineau, 1997; Krumhansl & Kessler, 1982; Parncutt & Bregman, 2000; Tillmann et al., 2008). These studies have generally found stronger effects of tonality on response time than studies that used single notes (Krumhansl & Shepard, 1979; Marmel et al., 2008; Warrier & Zatorre, 2002).

To address this concern, we used chords to provide a more robust and unambiguous establishment of tonal context and to remove the potential ambiguities of the contexts used in Experiment 4.1. We also redefined the No Context condition in Experiment 2 to include noise bursts preceding the test interval, in order to preserve attentional and temporal cuing without pitch reinforcement.

Since musically trained listeners are more sensitive to tonal hierarchies than listeners without musical training (Krumhansl & Shepard, 1979), any effect of context may be greater in musicians than non-musicians. Indeed, for relative pitch tasks, listeners with musical experience may be uniquely sensitive to preceding context that induces tonality (Dowling, 1986). Using the results from both Experiment 4.1 and Experiment 4.2, we also investigated whether participants with musical training were more likely to see an advantage from Major Scale context.

Method

Stimuli. All pitches were carried by harmonic complex tones, generated and presented in the same manner as in Experiment 4.1. In the No Context condition, noise bursts were generated with overall spectral shapes similar to those of the harmonic complex tones. Specifically, the bursts consisted of band-pass noise with a center frequency of 200 Hz and shallow filter slopes of 12 dB/octave. Like the harmonic complex tones, the noise bursts had durations of 400 ms, including 10-ms raised-cosine onset and offset ramps.

In Experiment 4.2, the harmonic context sequences were designed to establish a clear tonic, but without presenting the same pitch class twice, as otherwise participants could in theory compare the final tone in the test interval to a context tone one octave lower. This constraint led us to keep the context sequence very short. Figure 4.3 illustrates the design for harmonic context in each of the same five conditions. In the Noise Context condition, the test interval was preceded by two noise bursts. The F0 of the first test tone was randomly chosen from trial to trial from a range of 200 to 566 Hz with uniform distribution on a logarithm scale, just as in Experiment 1. In the Repetition condition, the test interval was preceded by two context tones with the same F0 as the first test tone. In the remaining three conditions, the context sequence consisted of two simultaneous tones followed by three simultaneous tones, and the F0 of the highest of the three simultaneous tones was always equal to the F0 of the first test tone. In the Mistuned Scale condition, the pitches in the context sequence were all related by multiples of 1.5 semitones. In the Whole-Tone Scale condition, they were all related by multiples of 2 semitones. In the Major Scale condition, the context sequence resembled an imperfect authentic cadence, establishing the pitch of the first test tone as tonic in a major key. It is important to note that the pitch distances are very similar between the final three conditions, although these small differences in pitch distance give rise to large differences in subjective sound quality of the resulting chords.

Participants. A new group of 20 participants, 6 male and 14 female, was recruited from the Twin Cities campus of the University of Minnesota. These participants ranged from 19 to 44 years of age (*Mean* = 24.4, *SD* = 6.4), and 0 to 35 years of reported musical experience (*Mean* = 6.0, *SD* = 9.0). They were all screened for normal hearing thresholds, as explained in Experiment 4.1. The experiment was conducted under the same IRB approval as Experiment 4.1, again in a single 2-hour session per participant. All participants provided written informed consent and were compensated $10 per hour for their participation.

Procedure. The basic task, to discriminate between "small, in-tune" and "large, mistuned" intervals, was the same as in Experiment 4.1. The one-second time limit on responding was again implemented, but again was triggered only rarely: in Experiment 4.2, the percentage of trials on which a participant ran out of time ranged from 0.1% to 6.3% (μ = 1.25%, *SD* = 1.65%).

For similar reasons (the novelty of the task for participants), participants again completed orientation and training before the testing phase. Orientation and training for Experiment 4.2 were slightly modified from Experiment 4.1. In the orientation phase, participants heard 5 labeled examples of small, in-tune and 5 examples of large, mistuned intervals. For this demonstration, $\Delta F0$ was fixed at 25.1%. The increase in the demonstration value of $\Delta F0$ (from 8% in Experiment 4.1) was to ensure that participants oriented quickly to the task.

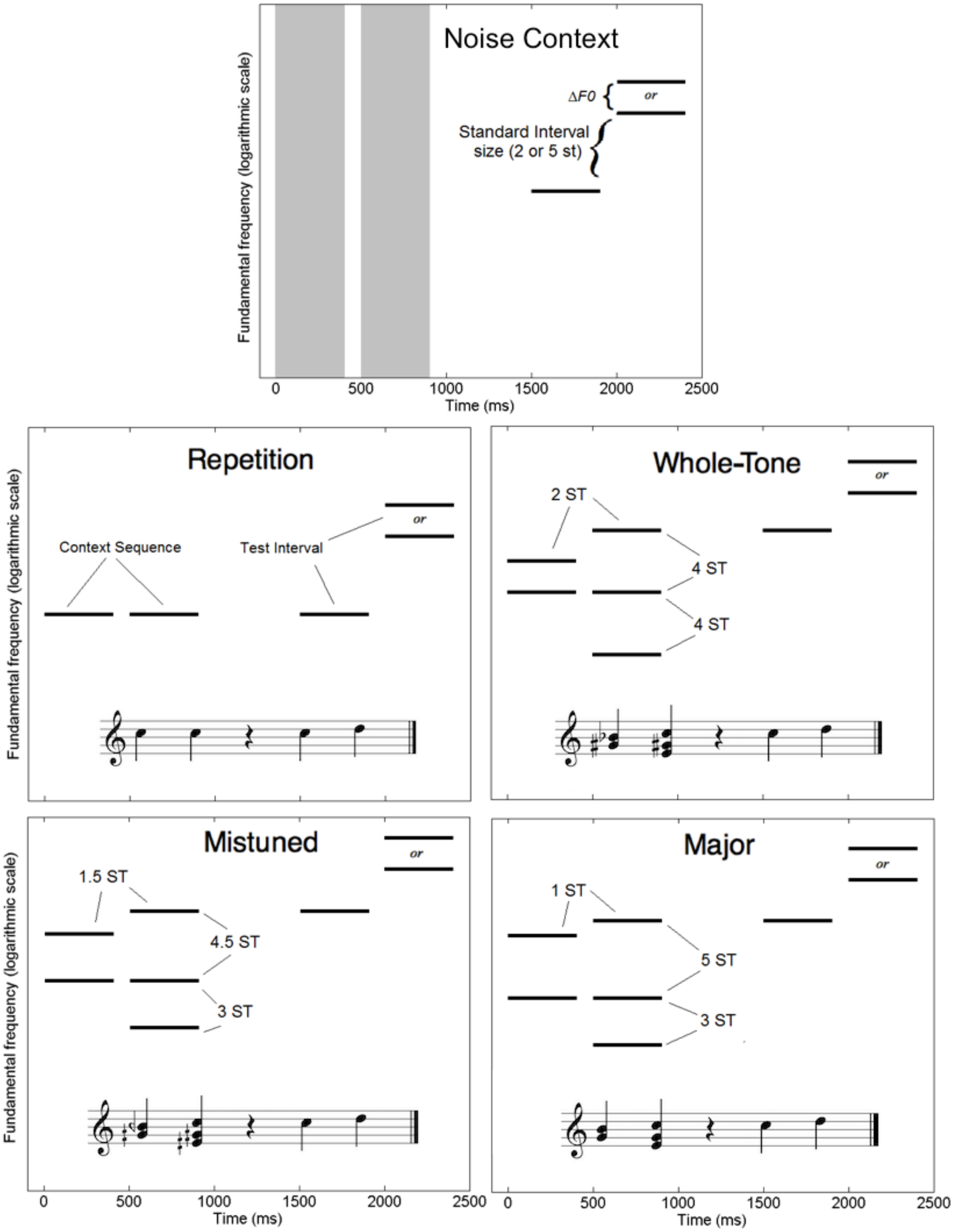


Figure 4.3. Schematic diagrams in spectrogram form of a single trial in each of the five context conditions for Experiment 2 (harmonic context). Pitch distances are labeled in semitones (ST). Gray vertical bars represent noise bursts. Context conditions are also illustrated with musical notation.

The training phase for Experiment 4.2 consisted of 4 blocks of 20 trials each in the Noise Context condition (test interval preceded by noise bursts). For the first block, $\Delta F0$ was fixed at 25.1%, for the second it was 15.8%, for the third 10%, and for the fourth 6.3%. These values of $\Delta F0$ are roughly in the region of one to four semitones. Once again, no time limit was imposed. As in Experiment 4.1, participants made few errors during this training phase, since the values of $\Delta F0$ were chosen to be easily discriminable, well above threshold.

An estimate of each participant's DL was obtained by presenting them with a range of values for $\Delta F0$ in the Noise Context condition and choosing the lowest level at which their performance fell between 60% and 85% correct. They were then tested at exactly this level in the five context conditions. The geometric mean estimated DL was 1.5% for the 2-semitone standard, 95% CI [0.9% 2.3%], and 1.2% for the 5-semitone standard, 95% CI [0.9% 1.7%]. These estimated DLs are much lower than those obtained in Experiment 4.1, likely due to the different measurement procedure, and pilot testing indicated that they were in the sensitive region (below ceiling and above chance) for later performance in the testing phase. Thus, unlike in Experiment 4.1, $\Delta F0$ testing levels were set equal to the estimated DLs in Experiment 4.2, not multiplied by 25.12%.

As in Experiment 4.1, the testing phase consisted of 25 blocks of 20 trials each (five blocks per condition, presented in pseudorandom order) and the order of the standard interval procedures was counterbalanced between participants, such that 11 participants completed all conditions with the 2-semitone standard first, while 9 participants completed all conditions with the procedure for the 5-semitone standard first.

Analysis. The sensitivity measure *d′* was calculated for each participant, standard interval size, and context condition, based on performance in the testing phase. The value of *d′* in the Noise Context condition was used as a baseline measure, which should have been roughly equal between participants and between conditions, based on the values of F0 chosen from the pilot phase of the experiment. Indeed, a paired-samples t-test on these *d′* values in the Noise Context condition revealed no significant difference between standard interval sizes (p = 0.43). Figure 4.4 summarizes performance in all conditions for both standard interval sizes. All five context conditions were analyzed with a repeated-measures ANOVA on *d′* values.

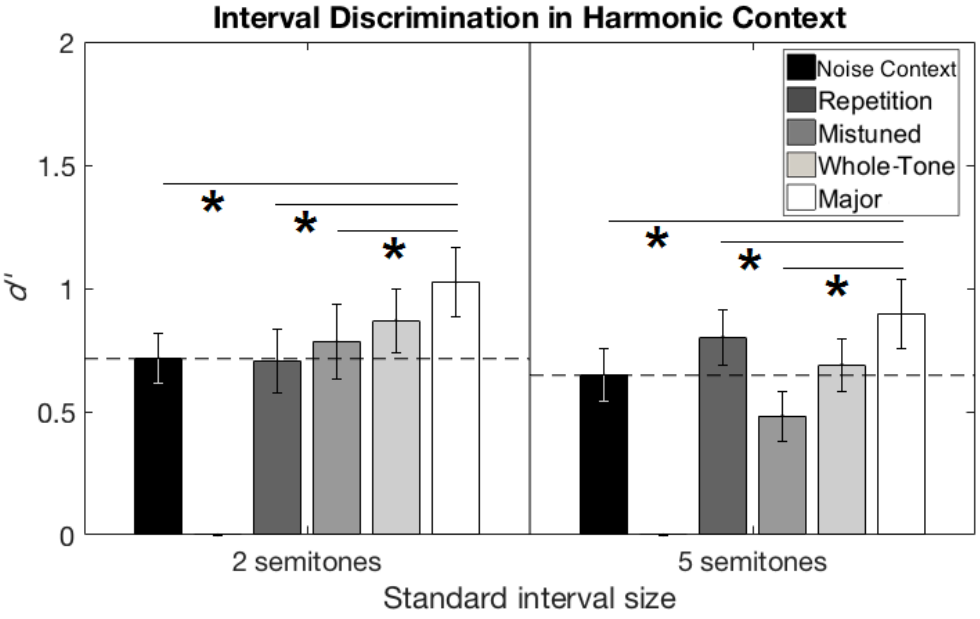


Figure 4.4. Interval discrimination performance from Experiment 4.2 (harmonic context). Interval discrimination performance in *d′* is shown for the 2-semitone (left) and 5-semitone (right) standard interval sizes. Performance in the Noise Context condition was treated as a baseline (horizontal dashed line). Error bars represent +/- 1 standard error of the mean across participants. Horizontal solid lines with asterisks show significant pairwise comparisons between conditions across both standard interval sizes.

Results

A repeated-measures ANOVA on *d′* values, considering the within-subjects factors of standard interval size and context condition, revealed an effect of context, *F*(4, 76) = 5.47, *p* = .001, η^2 = .224, but no other main effect or interaction. Pair-wise comparisons of context conditions with Bonferroni correction (criterion *p* = 0.05/10 = .005) showed an advantage for Major Scale over the baseline Noise Context (mean difference = .28, *p* = .003, Cohen's *d* = .59), as well as two of the other context conditions: Repetition (mean difference = .21, *p* = .003, Cohen's *d* = .45) and Mistuned (mean difference = .33, *p* = .003, Cohen's *d* = .65), with no other comparisons reaching significance.

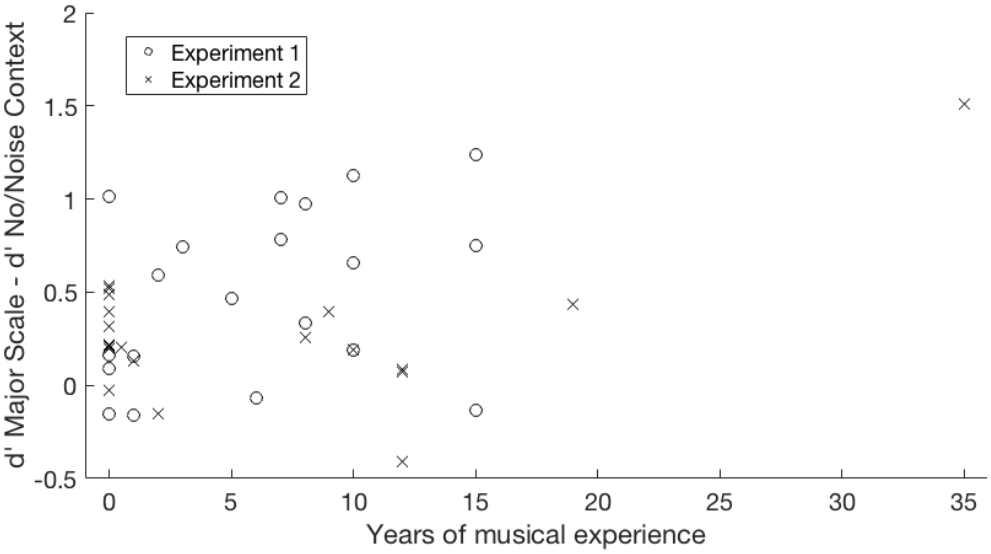


Figure 4.5. Benefit from familiar tonal context, measured as performance in the Major Scale condition minus performance in the No Context or Noise Context condition, as a function of years of musical experience. Data are plotted separately for Experiment 4.1 (melodic context, circles) and Experiment 4. 2 (harmonic context, X's).

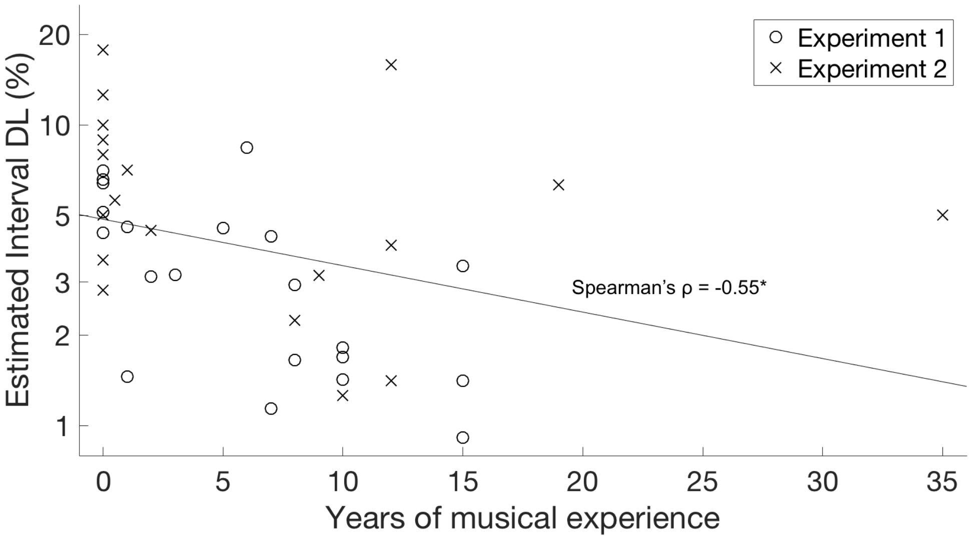


Figure 4.6. DLs for interval discrimination in No Context or Noise Context conditions as a function of years of musical experience. Data are plotted separately for Experiment 4.1 (melodic context, circles) and Experiment 4.2 (harmonic context, X's). The least-squares line across both experiments is plotted. A rank-order Spearman correlation was significant, *p* < .001.

For the three-way repeated-measures ANOVA including time, context, and standard interval size, the main effect of time was not significant *F*(1,39) = .60, *p* = .44, nor did any significant interaction involving time reach significance. Therefore, it seems that no substantial learning effects occurred over the course of the experiment.

A Spearman rank-order correlation revealed no significant relationship between years of musical experience and benefit over Major Scale relative to No Context, as shown in Figure 4.5 (Spearman's ρ = .17, *p* = .13). However, a significant negative correlation was observed between musical experience and baseline DL in the No Context condition, as shown in Figure 4.6 (Spearman's ρ = -.55, *p* < .001).

One participant reported 35 years of musical training, and thus could be considered an outlier (across both experiments, the mean amount of musical training was

5.9 years, and the standard deviation was 7.2 years). For all results in Experiment 4.2, inclusion or exclusion of this outlier had no effect on any of the statistical conclusions.

Discussion

The results of Experiment 4.2 were generally consistent with the hypothesis that establishing a familiar tonal context will improve the perception of pitch intervals; however, the observed benefit to performance was small in terms of changes in sensitivity index, *d′*, as in Experiment 1. In this experiment, the Major Scale context provided an advantage over all of the other context conditions except Whole-Tone Scale.

One of the goals of Experiment 4.2 was to discover whether a more strongly and unambiguously established tonal context would elicit a stronger effect. We did not find this result. The effect sizes from Experiment 4.2 are comparable to those from Experiment 4.1. Furthermore, in terms of the sensitivity index *d′*, the average benefit of the Major Scale context over No Context condition was somewhat lower in Experiment 4.2 (.28) than Experiment 4.1 (.47). Although these effects are statistically significant, a benefit of less than 0.5 in terms of *d′* represents a relatively small real-world advantage, considering this index ranges from 0 at chance performance to greater than 2.5 at ceiling (for 90% hits and 10% false alarms, *d′* = 2.56).

Experiment 4.2 was also designed to determine whether the removal of tonal ambiguities present in the melodic context of Experiment 4.1 would lead to a more definite advantage for Major Scale context over Repetition context. This was confirmed by the results, as performance in the Major Scale context in Experiment 4.2 was significantly better than performance in Repetition context. Again, however, this benefit was rather small, averaging only .21 d' units when combined across the two standard interval sizes.

It is necessary to acknowledge two possible alternative explanations for the smaller-than-expected benefit of tonal context in Experiment 4.2, when comparing the results to those of Experiment 1. In Experiment 4.2, the context sequences were shorter than in Experiment 1, consisting only of two sounds rather than four. It is possible that reducing the length of the context sequence reduced the overall benefit from any context condition, which may be why the benefit even from repetition context was smaller in Experiment 4.2. This possibility could be further explored with a longer harmonic context sequence. It is also possible that the introduction of noise bursts in the Noise Context condition aided performance in this condition compared to the No Context condition in Experiment 4.1, leading to less room for improvement with tonal context. Adding noise bursts to the No Context condition in Experiment 4.1 could potentially resolve this question. Both of these alternative explanations represent limitations in the validity of comparing results from Experiment 4.1 to results from Experiment 4.2, and should be kept in mind when making this comparison.

IV. General discussion

The present study investigated the effect of familiar tonal context on the perception of pitch intervals. We hypothesized that when discriminating between a small, in-tune interval and a large, mistuned interval, participants would perform better with an established (major) tonal context than with no context, or other less well-established contexts. The main results were that the Major Scale conditions provided an advantage over No Context in both experiments, over the Whole Tone condition in Experiment 4.1, and over the Repetition and Mistuned conditions in Experiment 4.2. An interaction in Experiment 4.1 revealed different results depending on standard interval size, with Major Scale only providing an advantage over No Context and Whole conditions for the 2-semitone standard interval size.

Although we observed an effect of tonal context on interval perception, and a particular benefit for familiar tonal context, these conditions provided only a small benefit to performance over the No Context and Noise Context conditions, in terms of *d′*. These results suggest that learned tonal hierarchies may influence the accuracy of pitch interval perception, although the benefit may be slight in practical terms.

Using nonparametric Spearman correlation, we observed no significant correlation between the benefit of Major Scale context over No Context and years of musical experience. This outcome suggests that the benefit of familiar tonal context was not dependent on amount of musical training. In this respect, the outcomes are consistent with the conclusion of Bigand and Pineau (1997) that non-musicians are as sensitive to tonal structures as musicians, as well as more recent findings showing that when judging singers, even non-musicians are highly sensitive to mistunings from the equal-temperament scale (Hutchins, Roquet, & Peretz, 2012; Larrouy-Maestri, Magis, Grabenhorst, & Morsomme, 2015), presumably learned through passive exposure. It is also consistent with the finding that even non-musicians exhibit early right-anterior negativity (ERAN) in response to violations of tonal expectations (Koelsch, Gunter, Friederici, & Schröger, 2000). Importantly, however, this outcome does not imply that musicians and nonmusicians were equally sensitive in general to pitch intervals, firstly as all scores are normalized by performance in the no/noise context condition, and secondly as all participants were tested at their individually estimated DL, which varied between participants. Indeed, the significant correlation between DLs and musical experience indicates that musically trained listeners had higher baseline performance on this task in the No Context condition.

It is worth noting that the task used in this study, in which participants compare a small, in-tune interval to a larger, mistuned interval, differs from that used in previous studies. The goal was to design a task that could in theory be strongly affected by tonal context. This task must be done based on the sizes of the intervals unless participants have access to tuning cues, which should in theory make the task much easier with tonal context. Another advantage of this task is that it allows for a shorter time between the end of the context sequence and the end of the test stimulus, also theoretically maximizing the effect of context. For these reasons, it seems likely that other tasks (removing the tuning cue, or traditional two-interval comparison) would likely find even smaller context effects. However, the conflation of tuning and interval size can be considered a limitation of the present study, in that performance may reflect access to tuning cues, precision of perceptual representation of interval size, or both. A task with no tuning cue (for example, with both intervals equally spaced around a “standard” common Western tonal interval) could be used to better resolve this question.

Our ability to discriminate very small changes in F0 on basic discrimination tasks is not mirrored in interval discrimination measured in isolation, where thresholds can be considerably higher (e.g. McDermott, Keebler, et al., 2010). One possibility explored by the present study was that tonal context would play a much larger role in musical interval size discrimination. We had hypothesized that tonal hierarchies, by allowing listeners to hear successive pitches as points within a structure rather than as successive interval sizes, could provide a context in which performance for interval discrimination might better approximate performance for basic discrimination. In the present study, performance on interval discrimination was improved by the presence of a tonal context that suggested a familiar tonal hierarchy (major scale) to a greater degree than the improvement from the presence of incongruent tonal context. However, the degree of benefit was small, and was not sufficient to account for the differences in DLs between basic discrimination and interval discrimination, suggesting that even with tonal context, interval size perception is not as precise as basic pitch perception.

When cued using harmonic progressions, tonal context has robust effects on pitch processing as measured by response time (Bigand & Pineau, 1997; Tillmann et al., 2008). As discussed in those studies, the increase in response time in the presence of an unexpected or incongruous chord likely reflects cognitive priming, where the unexpected chord interferes with the detection of the target mistuning, rather than any enhancement of pitch representation through the tonal context in congruent conditions. The results from our study provide support for this interpretation: pitch interval discrimination is barely affected by congruent tonal contexts relative to priming by a single tone, and not affected at all in melodic context, where the benefit of each of these conditions relative to no context is dependent on standard interval size. One interpretation of our outcomes is that both harmonic and melodic context were effective in establishing a tonal hierarchy, but that tonal context provides only a modest benefit to pitch interval perception. Another possibility is that even the harmonic context sequences were not sufficiently long or harmonically rich to fully establish a musical key, such as the longer chord progressions used in previous studies.

The inclusion of control conditions in our study allowed us to further examine the source of the benefit from musical context to interval discrimination. In an earlier study of interval categorization and labeling in musicians (Wapnick et al., 1982), melodic context provided a large benefit over interval perception in isolation. However, our results suggest that some of this benefit may have merely been due to an effect of pitch reinforcement, as can be seen in the observed benefit from the Repetition context condition in Experiment 4.1 (melodic context). We can also conclude, however, that tonal congruence is a necessary component of effective tonal context, because context that established a familiar major-key tonality was superior to context conditions that failed to establish this tonality, such as the Mistuned Scale (in Experiment 4.2) and Whole-Tone Scale (in Experiment 4.1) contexts. For the harmonic (chord) contexts in Experiment 4.2, the tonally congruent context was also more effective than simple pitch reinforcement through repetition.

We measured interval discrimination thresholds at two standard interval sizes, and, consistent with previous findings, we found no significant effect of standard interval size on discrimination thresholds (Burns & Ward, 1978; McDermott, Keebler, et al., 2010). However, standard interval size did have an effect on the pattern of results in the various context conditions in Experiment 4.1. Specifically, for a 2-semitone standard interval size, the best performance was from the Major Scale condition, whereas for the 5-semitone standard interval size, the best performance was from the Repetition condition. This difference may be driven in part by the different functions of the corresponding scale degrees in the major musical key where the reference tone is tonic. Both the major second (2 semitones) and the perfect fourth (5 semitones) are within the major key, but the perfect fourth is especially stable and closely related to tonic (Krumhansl & Kessler, 1982). The two next-largest intervals to these, against which participants would have to discriminate in the present study if $\Delta F0$ was near 1 semitone, also have different functions: neither the minor third (3 semitones) nor the tritone (6 semitones) belong to the major key, but the minor third is traditionally thought of as a consonant interval in Western music, while the tritone is thought of as dissonant.

Given the clear difference in tonal belongingness and consonance between the perfect fourth and the tritone, participants' relatively poor performance in the 5-semitone standard task –even with tonal context – is surprising. One relevant factor may be that all intervals in the present study were defined using the equal-temperament tuning system, wherein 5 semitones is defined as $2^{(5/12)}$. While some evidence suggests that a semitone defined in equal-temperament terms may be a perceptually relevant boundary for interval discrimination (Zarate, Ritson, & Poeppel, 2012), human ideals for musical intervals, whether measured by listener adjustment or musician production, do not agree perfectly with equal-temperament tuning, and this discrepancy may be even greater when measured within a musical context (Rakowski, 1990). Specifically, it appears that musical context may actually increase the deviation of this "ideal" from an equal-temperament standard for the ascending tritone. If this is the case in our experiment, participants may have greater difficulty discriminating an ascending perfect fourth from an ascending tritone (both defined by equal-temperament tuning) if their internal tuning allows for the tritone to be larger. In other words, it is possible that, with $\Delta F0$ generally less than 1 semitone, both the “small, in tune” and “large, mistuned” intervals in the 5-semitone-standard task would fall into the “perfect fourth” category of musicians’ categorical interval perception, rather than straddling the boundary between the “perfect fourth” and “tritone” categories.

In summary, we explored the effect of familiar tonal context on a pitch interval discrimination task, a performance-based measure of pitch interval perception accuracy. In contrast to expectations of strong effects of tonal context on the accuracy of pitch-interval discrimination, we found a relatively small benefit. The results suggest that although tonal contexts can generate strong expectations, they do not produce substantial enhancements in the perceptual representations of pitch and pitch intervals.

Chapter 5: Simultaneous pitches: behavioral results for three concurrent pitches

I. Introduction to multiple complex pitch perception

Human listeners are generally able to perceive multiple pitches at the same time without great difficulty. In music, the presence of three or more concurrent pitches is the rule, not the exception. However, although a great deal of psychoacoustic research has been conducted on pitch perception for single harmonic complexes, and some on the perception of two-complex mixtures (e.g. Beerends & Houtsma, 1989; Carlyon, 1996; Micheyl, Bernstein, & Oxenham, 2006; Micheyl, Keebler, & Oxenham, 2010; Wang et al., 2012), mixtures of three or more concurrent complexes have not received as much attention. Furthermore, while humans regularly identify multiple simultaneous pitches occurring in music, existing computational algorithms for the estimation of a single pitch (de Cheveigné & Kawahara, 2002; Noll, 1967) are far more effective than algorithms estimating multiple simultaneous pitches (de Cheveigné & Kawahara, 1999; Klapuri, 2008; Yeh et al., 2010). Besides being ecologically valid, mixtures of three or more concurrent complexes can provide a strong test for models of pitch perception, because adding more complexes with different fundamental frequencies (F0s) greatly decreases the peripheral resolvability of components in the resulting mixture, but does not increase harmonic numbers of the components present in the mixture. This dissociation of harmonic number and peripheral resolvability is important because it leads to different predictions from different models of pitch perception. Some models (e.g. Larsen, Cedolin, & Delgutte, 2008) rely on resolvability after mixing, working well only when individual components can be isolated in the rate-place representation. Other models (e.g. Bernstein & Oxenham, 2005) depend instead on harmonic number before mixing, working well regardless of resolvability, as long as F0 is not far from CF, and some included components are below the limit of phase locking.

The pitch of a single missing-F0 harmonic complex significantly changes in quality when components below about the 10th are removed, a phenomenon for which the most likely explanation is the decreased peripheral resolvability of high-numbered harmonics. The empirical hallmarks of pitch from unresolved harmonics beyond about the 10th component are elevated F0 difference limens (DLs) and sensitivity to additive phase relations, suggesting sensitivity to temporal interactions only between high-numbered components (Houtsma & Smurzynski, 1990). Doubling the temporal envelope repetition rate by adding components in alternating sine-cosine phase doubles perceived pitch only for high-harmonic complexes (Shackleton & Carlyon, 1994). Peripheral resolvability is also implicated by findings that the transition point between normal and elevated F0DLs shifts downward from the 10th component in situations where auditory filter bandwidths are broadened. Specifically, this shifted transition point is observed for high-level stimuli (Bernstein & Oxenham, 2006b), in listeners with sensorineural hearing loss (Bernstein & Oxenham, 2006a), and in age-related hearing loss (Patterson et al., 1982; Russo et al., 2012).

The importance of peripheral resolvability to F0DLs seems to suggest a rate-place explanation: the shift in F0DLs can be explained by a shift in degree of peripheral resolvability (as measured based on filter bandwidths, e.g. Glasberg & Moore, 1990). However, unresolved harmonics, which provide no rate-place information, still produce a pitch sensation that can be discriminated (albeit at a higher F0DL). Temporal models (e.g. Meddis & O'Mard, 1997) can explain pitch for unresolved harmonics as phase-locking to the temporal envelope (the overall period when multiple components sum and interact in one filter), whereas pitch from resolved harmonics could be coded by phase-locking to TFS (the period of each individual component). Most temporal models do not explain why pitch should be more salient from resolved than unresolved harmonics, but some autocorrelation models have been adapted to do so, e.g. by introducing a place-dependent lag window (Bernstein & Oxenham, 2005), or by using synthesized delays with maximum durations limited by cochlear filter bandwidth (de Cheveigné & Pressnitzer, 2006). Thus both rate-place and adjusted temporal models are capable of explaining lower F0DLs for resolved harmonics, but temporal models do so by positing that harmonic number, not resolvability, is responsible.

One way to dissociate resolvability from harmonic number is by simultaneously presenting a second, spectrally overlapping harmonic complex with a different F0. This decreases the resolvability of the components in each complex, but does not change harmonic number. Research using these kinds of stimuli has provided conflicting evidence about the importance of resolvability vs. harmonic number. Presenting a complex dichotically, such that the odd harmonics are presented to one ear and the even harmonics to the other ear, improves resolvability but does not strengthen pitch perception as measured by F0DLs (Bernstein & Oxenham, 2003). In a complementary finding, slightly mistuning the odd harmonics of a complex relative to the even harmonics (by 3%) improved F0DLs without improving resolvability (Bernstein & Oxenham, 2008). Both of these findings are better explained as effects of harmonic number than of peripheral resolvability. However, other studies using two overlapping complexes point to resolvability. For pitch discrimination of a target complex tone in the presence of a masking complex tone, target-to-masker ratio (TMR) thresholds are lowered by perceptual segregation only when resolved harmonics are present after mixing (Micheyl, Bernstein, et al., 2006), and F0DLs are low only when salient (> 1 dB) peaks are present in modeled excitation patterns (Micheyl et al., 2010). If harmonic complexes are unresolved even before mixing, no pitch is perceived (Carlyon, 1996; Micheyl et al., 2010), while for a target complex tone with marginally resolved harmonics before mixing, an effect of additive phase relations also appears when a masker complex tone is concurrently presented, thus decreasing resolvability (Wang et al., 2012). Performance on a pitch identification task for two concurrent two-tone complexes is degraded when they are presented diotically instead of dichotically, but only in conditions where none of the four partials is individually resolved (Beerends & Houtsma, 1989).

Different means of dissociating resolvability and harmonic number have thus produced different answers to the question of which drives the transition between normal and elevated F0DLs. Mistuning and dichotic separation suggest that harmonic number is more important, while inclusion of a concurrent masking complex suggests resolvability is more important. In an effort to resolve this conflict, we used stimuli that provide an even stronger degree of dissociation between harmonic number and resolvability: mixtures of three concurrent harmonic complexes. We measured pitch discrimination of tones within these mixtures with three different behavioral tasks in experiments 5.1-5.3, in the hopes of providing a more definitive answer to the question of whether resolvability or harmonic number is responsible for high-harmonic weak pitch.

II. Experiment 5.1: Major and minor triad discrimination

Rationale

The goal of this experiment was to test whether listeners can perceive three simultaneous pitches to a precision of one semitone (typically near the DL for unresolved pitch, e.g. Houtsma & Smurzynski, 1990) in a three-complex mixture, in conditions where the individual complexes contain resolved harmonics before mixing (and thus include low-numbered harmonics), but the mixture contains no resolved harmonics. If resolvability is the determining factor in pitch strength, the task should become impossible in conditions where the mixture contains no resolved harmonics. On the other hand, if harmonic number is the determining factor, performance should remain high as long as low-numbered harmonics are included, even if the mixture does not contain resolved harmonics.

Methods

Listeners. Thirty normal-hearing listeners (21 female and 9 male) were initially recruited for the experiment, ranging in age from 19 to 44 (mean = 24.93), and in years of musical experience from 0 to 35 (mean = 9.03). All listeners had audiometric thresholds of 20 dB HL at octave frequencies between 250 and 8000 Hz. Out of the recruited 30 listeners, only 9 passed the initial pure-tone screening task, and these 9 continued on to complete the complex-tone task and F0DL measurements. These 9 participants, 6 female and 3 male, ranged in age from 22 to 30 (mean = 26.1), and in years of musical experience from 2 to 19 (mean = 10.4).

Stimuli. We presented listeners with major and minor triads in root position, 1st inversion, and 2nd inversion. In Western music theory, these are combinations of three pitches that are separated from each other by specific numbers of semitones (ST), where a semitone is defined in the equal-temperament tuning system as a ratio of $2^{1/12}$ between F0s, such that an octave (doubling of F0) is divided equally into 12 semitones. In any of its three inversions, a major triad is always only one ST away from a minor triad, and vice versa. In root position, the pitches of a major triad are 0, 4, and 7 ST above the lowest pitch, while the corresponding pattern for a minor triad is 0, 3, 7. In 1st inversion, major is 0, 3, 8, and minor is 0, 4, 9. In 2nd inversion, major is 0, 5, 9, and minor is 0, 5, 8. In Western music, these two chords are common and the distinction between them is important. On each trial, we presented listeners with one triad and asked them to indicate whether it was major or minor. Inversion (root, 1st , or 2nd) varied randomly from trial to trial, and the F0 of the lowest tone was roved between 200-230 Hz, so listeners had to perceive all three pitches and their relationship to each other in order to successfully perform the task. Importantly, perceiving any two of the three pitches is not sufficient to perform this task: for example, even after determining that the upper two pitches are 3 semitones apart, a listener must also hear the third (lowest) pitch to know whether the chord is major, root position (0, 4, 7) or minor, 2nd inversion (0, 5, 8). This is true for any combination of two out of the three pitches, so long as all three inversions are possible and the absolute pitch range is variable.

For the pure-tone screening task, each trial consisted of three concurrent 500-ms pure tones whose frequencies formed either a major or minor triad, with all frequencies between 200 and 387 Hz. Each pure tone had a level of 50 dB SPL, with 10-ms raised-cosine on- and off-ramps.

For the complex tone task, each trial consisted of three concurrent 500-ms harmonic complex tones whose F0s formed a major or minor triad, with all F0s between 200 and 387 Hz, filtered into one of four different spectral regions, with low cutoffs of 0.5, 2, 3, and 4 kHz. The high cutoff was 2 kHz for the first spectral region and 8 kHz for the rest. Figure 5.1 shows spectra of single tones and triads in these four spectral regions. Filter slopes were 24 dB per octave on either side.

The spectral regions were chosen to provide different degrees of harmonic resolvability (Houtsma & Smurzynski, 1990). In the lowest region, resolved components were always present, with the lowest component in the passband ranging from the 2nd to the 3rd harmonic depending on F0. In the highest region, the individual complexes were unlikely to contain resolved harmonics even before mixing, with the lowest component in the passband ranging from the 11th to the 20th harmonic, depending on F0. In the middle two regions, harmonic number and resolvability were more dissociated. The complexes included low-numbered harmonics and would have likely been resolved prior to mixing, with lowest components ranging from the 6th to the 10th in the 2-8 kHz region, and from the 8th to the 15th in the 3-8 kHz region. However, given that the introduction of two concurrent complex tones decreases the averaging spacing between components by a factor of 3, the resulting mixture of three complexes was unlikely to contain peripherally resolved harmonics.

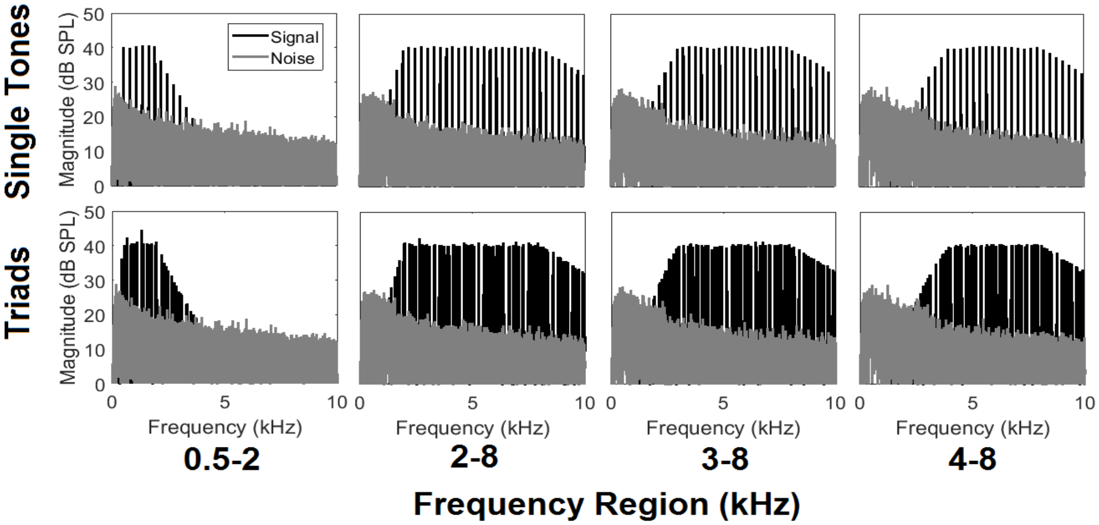


Figure 5.1. Frequency spectra of example stimuli used in Experiments 5.1-5.3, embedded in threshold-equalizing noise (TEN) presented to mask distortion products. The top row shows single complex tones with F0s of 275 Hz. The bottom row shows major root-position triads with F0s of 218, 275, and 327 Hz.

For F0DL measurements, each trial consisted of two 200-ms harmonic complex tones, separated by a 100-ms gap, with F0s geometrically centered around 260 Hz. The difference in F0 between the two tones changed adaptively from trial to trial. Tones were filtered into the same four spectral regions as in the complex tone task.

For both the complex tone triad task and the F0DL measurements, all complexes were presented at 40 dB SPL per component within the passband, and embedded in threshold-equalizing noise (TEN) (Moore, Huss, Vickers, Glasberg, & Alcántara, 2000) at 30 dB SPL within the ERB centered around 1 kHz. The noise started 300 ms before tone onset, and ended 200 ms after tone offset. All stimuli had 10-ms raised-cosine on- and off-ramps. In the F0DL task, the components were added in either sine (zero) starting phase or random phase, with the starting phase of each component selected at random from between 0 and 2π independently and with uniform distribution on each presentation.

All stimuli were generated within MATLAB (The Mathworks, Natick, MA), using a 24-bit L22 soundcard (LynxStudio, Costa Mesa, CA), and were presented diotically through HD650 headphones (Sennheiser USA, Old Lyme, CT) at a sampling rate of 44.1 kHz.

Procedure. Before anything else, listeners performed a pure-tone screening task, in which they distinguished major from minor triads, defined using pure tones at the F0s. The criterion to pass this screening was a performance level of ≥ 80 % correct for each inversion (root, 1st, 2nd) after a maximum of 5 training blocks, with each block containing 60 trials (20 trials per inversion). The purpose of this training/screening task was to ensure that listeners could reliably tell a major chord from a minor chord in simple conditions. Listeners who passed the pure-tone screening then completed the same task using missing-F0 complex tones filtered into four spectral regions (see "Stimuli"). Each listener completed a total of 100 trials per inversion in each spectral region. Inversion varied from trial to trial, and spectral region changed every block of 60 trials, with block order randomized. Finally, listeners' F0DLs for individual missing-F0 complexes were measured in the same four spectral regions, using a 1-up 2-down adaptive-tracking procedure, with each DL the average of 3 runs, and each run the average of 6 reversals.

Results

The results of the pure-tone screening task are shown in in the left panel of Fig. 5.2. All other results in Fig. 5.2 and this section are only for the 9 listeners who met the performance criterion of 80% correct in all 3 inversions within 5 blocks of pure-tone training.

As shown in the middle panel of Fig. 5.2, average F0DLs were below 1% in the lower two spectral regions, but increased in the higher two spectral regions, with DLs approaching one semitone (~6%) in the highest region, where an effect of component phase (sine vs. random) was also clearest. An ANOVA on the log-transformed F0DLs found a main effect of phase [F(1,8) = 12.8, $p = 0.007$], a main effect of spectral region [F(3,8) = 85.3, $p < 0.001$], and an interaction [F(3,24) = 4.97, $p = 0.008$]. Paired comparisons showed that the effect of phase was significant only for the highest spectral region ($p < 0.001$). The elevated DLs and the emergence of a phase effect in the higher spectral regions confirm that the single complexes in the higher two regions may not contain resolved harmonics.

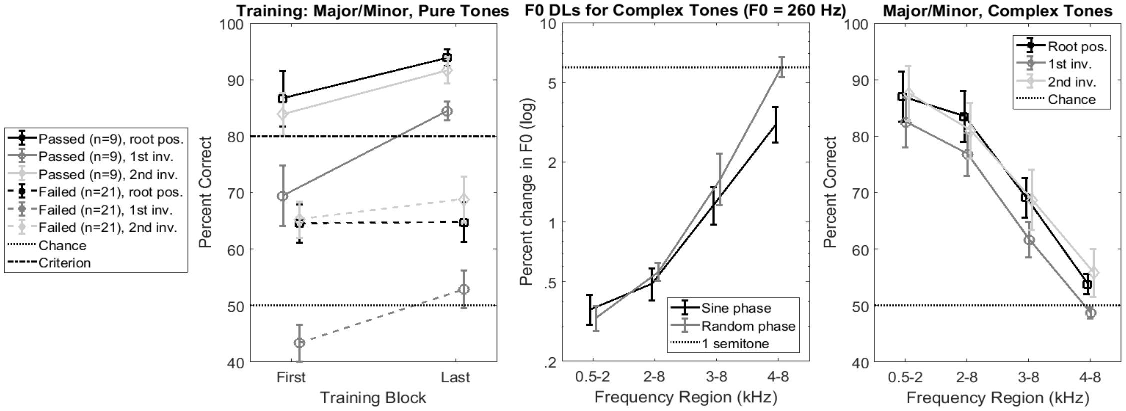


Figure 5.2. Results from Experiment 5.1. Left: results from pure-tone training on the major/minor discrimination task, for the 9 listeners who met the criterion and the 21 listeners who did not meet the criterion. Only the 9 listeners who passed went on to complete the rest of the experiment. Center: F0DLs for single complex tones in four spectral regions, with components added in sine or random phase. Right: results from the major/minor discrimination task with complex tones. Error bars show 1 SEM.

The results from the major-minor discrimination task are shown in the right panel of Fig. 5.2. An ANOVA on percent correct scores, transformed into rationalized arcsine units (RAU; Studebaker, 1985), found a main effect of spectral region [F(3,24) = 33.2, *p*

< 0.001], and a main effect of inversion [F(2,16) = 5.7, $p = 0.01$], but no interaction [F(6,48) = 0.3, $p = 0.93$].

Discussion

The pattern of results was similar for both the single-complex and three-complex experiments: for the single-complex experiments, F0DLs were low and, for random-phase harmonic complexes, only approached one semitone in the highest spectral region, where none of the harmonics would be considered resolved; for the three-complex experiment, performance was high and only decreased to near chance (reflecting an inability to discriminate a 1-semitone difference) in the same highest-frequency condition. Most importantly, performance in the three-complex experiment was high in the two middle spectral conditions, where the harmonics were likely unresolved in the mixture. Therefore, the results of Experiment 1 suggest that resolved harmonics might not be strictly necessary for accurate musical pitch perception.

III. Experiment 5.2: Hearing out one complex in the presence of others

Rationale

The results of Experiment 5.1 support the idea that low harmonic numbers, not necessarily resolved harmonics, are necessary for accurate pitch perception. However, it is possible that listeners may have been listening for some emergent holistic cue that distinguishes major from minor triads, rather than explicitly coding all three pitches in the combination. This seems unlikely, given that chord inversion (specific pitch relationships) was randomized from trial to trial and absolute F0 was also roved. Taken together, these two randomizations would have made any individual F0, or even two F0s, an unreliable cue. Due to the same two factors, patterns of fluctuations in the temporal envelope, caused by beating between nearby components, would also have been inconsistent. Thus, there are no obvious cues on which any putative holistic quality signaling major or minor could be based. Nevertheless, given this potential concern, along with the fact that only a third of the screened listeners were able to perform the task satisfactorily even with pure tones, this second experiment was run. In this experiment, we used a more explicit test of pitch resolution for the middle-F0 complex in a combination of three simultaneous complexes, by adapting a paradigm that has been used previously to study pure-tone pitch perception (Laurent Demany, Semal, & Pressnitzer, 2011).

Methods

Listeners. Twenty-two normal-hearing listeners participated in the experiment. Out of the recruited 22 listeners, 15 passed the initial pure-tone screening task, and these 15 continued on to complete the single-pitch and multiple-pitch complex-tone tasks. These 15 participants, 10 female and 5 male, ranged in age from 18 to 26 (mean = 20.7), and in years of musical experience from 0 to 19 (mean = 5).

Stimuli. The paradigm for Experiment 5.2 is illustrated on the left side of Figure 5.3. On each trial, listeners heard a single 300-ms reference tone, followed by a 600-ms gap, followed by a second 300-ms target tone (pure-tone screening and single-pitch condition) or a combination of three simultaneous 300-ms tones with different F0s (multiple-pitch condition), where the middle tone was the target and the higher and lower tones were maskers. The F0s of the reference and target tone were either 1 or 0.5 semitones apart, and the direction of the pitch change was either up or down. Listeners had to identify the direction of this pitch change. The low masker's F0 was roved between 200 and 224 Hz.

The reference tone and target tone F0s were geometrically centered on a nominal F0 that was roved from 3-6 semitones above the low masker's F0, so always fell between 238 and 317 Hz. The high masker's F0 was roved from 3-7 semitones above the nominal F0, so always fell between 283 and 475 Hz. All complex tones were filtered and embedded in TEN as in Experiment 5.1.

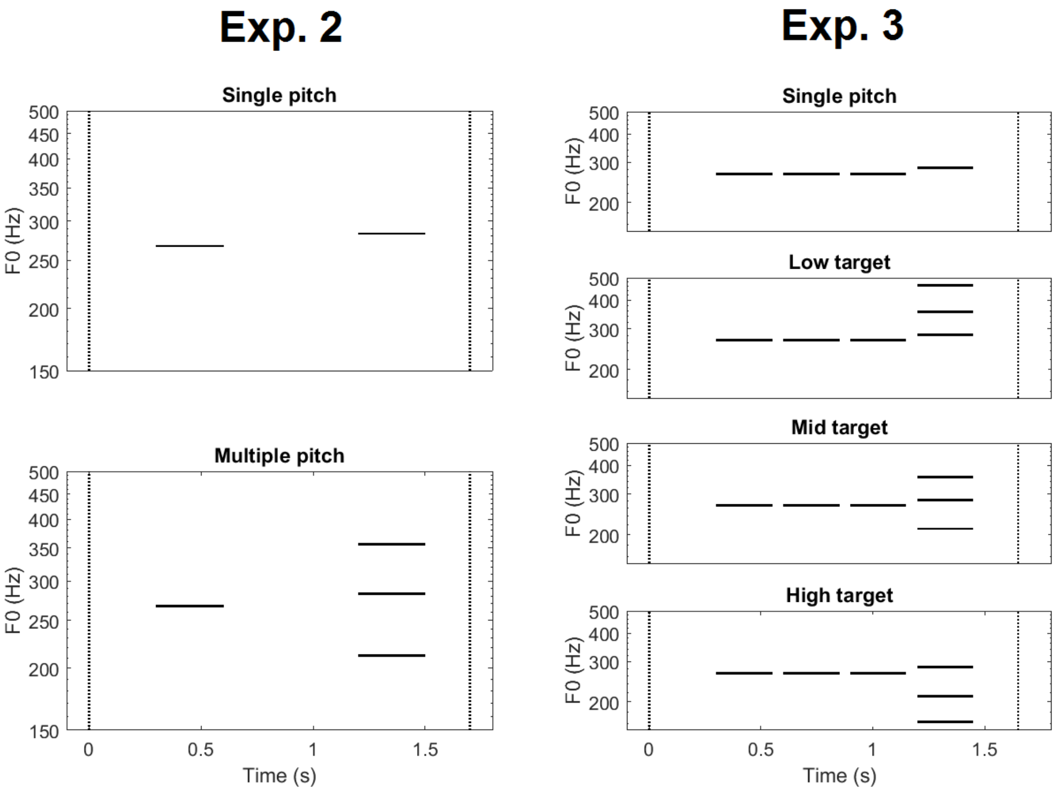


Figure 5.3. Example trials from experiments 5.2 and 5.3. Listeners identified the direction of a pitch change between reference tone(s) and target tone, with the distance between the two tones held constant at either 0.5 ST or 1 ST. The target tone was either presented alone, or as the lowest, middle, or highest of a three-tone mixture. Dashed vertical lines show onset and offset of TEN.

Procedure. In the same spectral regions used in Experiment 5.1, with missing-F0 complexes with similar F0s to Experiment 5.1, listeners in Experiment 5.2 identified the direction of a pitch change of 1 or 0.5 semitones between two complex tones, when the second tone is presented concurrently with two other complex tones, one with a higher F0 and one with a lower F0. If harmonic number truly drives elevated DLs more than resolvability, performance on this multiple-pitch task should be above chance in the same conditions as the major/minor discrimination task in Experiment 5.1. Listeners completed 100 trials total per F0 difference in each spectral region. F0 difference varied (either 0.5 or 1 ST) from trial to trial, and spectral region changed every block of 40 trials, with block order randomized. As a control condition, listeners' performance was also measured in a single-pitch condition where the second tone was also presented alone, with no concurrent masking complex tones. Listeners were screened with a task equivalent to the single-pitch condition using pure tones before advancing to other tasks, to ensure they could reliably perform the task in simple conditions. The criterion for the pure-tone screening task was 80% correct in both the 1 and 0.5 ST conditions, after a maximum of 3 training blocks, each of 100 trials per condition.

Results

The results from Experiment 5.2 are shown in Fig. 5.4. On their final training block, the 15 listeners who passed the pure-tone screening task responded correctly on an average of 98.9% of trials in the 1-ST condition, and 94.9% in the 0.5-ST condition. For the single-pitch task, performance deteriorated in the two highest spectral regions, where no harmonics below the 10th were present in the passband for at least some (3-8 kHz condition) or all (4-8 kHz) trials. However, performance remained above chance for all spectral regions, for both the 1- and 0.5-ST F0 difference between the reference and target tone. For the multiple-pitch task, performance was above chance in lowest 3 spectral regions, for both F0 differences. An ANOVA on percent correct scores, transformed into RAU, found a main effect of spectral region [F(3,42) = 81.7, $p < 0.001$], a main effect of task (single vs. multiple) [F(1,14) = 169.2, $p < 0.001$], a main effect of F0 difference (0.5 vs. 1 ST) [F(1,14) = 37.9, $p < 0.001$], and an interaction of task by spectral region [F(3,42) = 9.74, $p < 0.001$], but no other interactions. Pairwise comparisons revealed that all spectral regions differed significantly from each other ($p <$ 0.001 in all cases). Post-hoc comparisons investigating the task by spectral region interaction revealed that all spectral regions differed significantly from each other within each task ($p < 0.001$ in all cases) except for the lowest two spectral regions in the single-pitch task ($p = 0.18$).

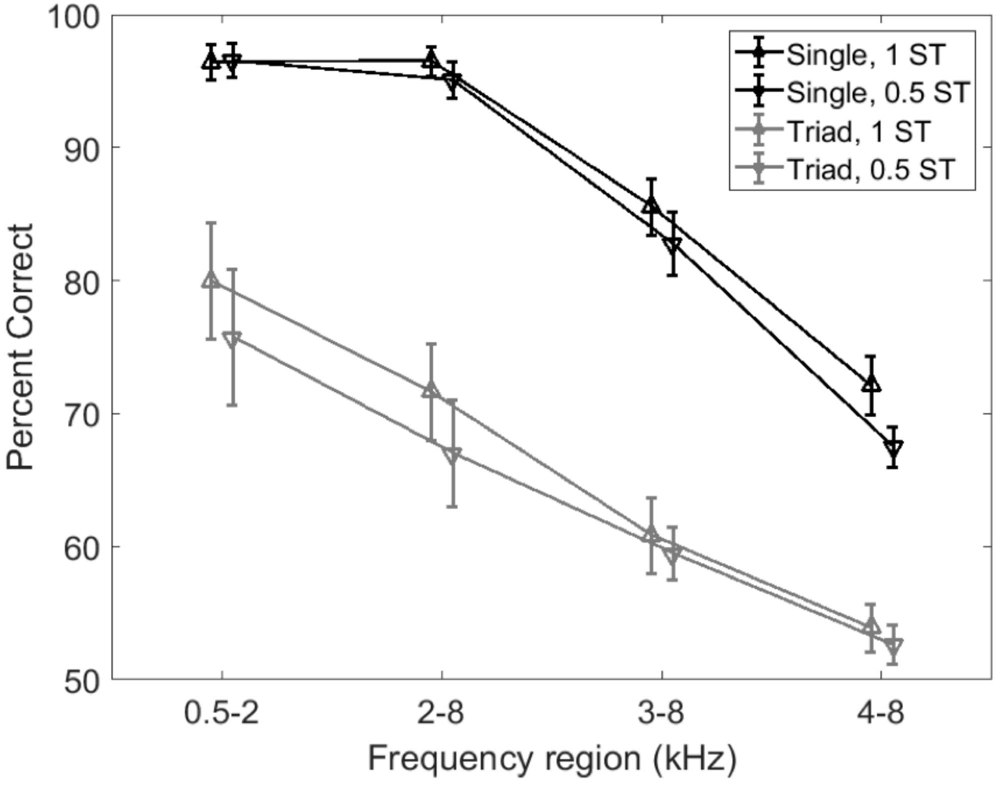


Figure 5.4. Results for Experiment 5.2. Listeners identified the direction of a pitch change in single-complex-tone and multiple-complex-tone conditions, with a constant F0 difference of either 1 or 0.5 semitones. Target F0 was roved between 238 and 317 Hz. Error bars show 1 SEM.

Discussion

The results suggest that listeners were able to discriminate a 0.5-ST F0 difference between two complexes, even they were filtered into a spectral region where the mixture was unlikely to contain any spectrally resolved harmonics. However, overall performance was worse in the multiple-pitch task than the in single-pitch task. One interpretation of the overall degradation in performance for the combination task relative to the single task is as an attention-related difficulty occurring centrally; this would be a similar effect to that observed by Beerends & Houtsma (1986) in their tests of dichotically presented simultaneous two-tone complexes. Another interpretation is that this difference is an effect of decreased resolvability, with participants performing worse for the combination task because resolvability is decreased there relative to the single task. In order to evaluate these two possible interpretations, in Experiment 5.3 we modified the paradigm to possibly facilitate attention to the target tone, and we also introduced a control condition with pure tones, to evaluate the gap between single-pitch and multiple-pitch performance when all stimuli are clearly resolved.

IV. Experiment 5.3: Hearing out the low, middle or high voice.

Rationale

The middle voices in polyphonic music are not typically as easy to follow as the outer voices, particularly the highest voice (Laurel J. Trainor, Marie, Bruce, & Bidelman, 2014). Since listeners in Experiment 5.2 were tasked with hearing out the middle voice out of three, it seems possible that performance in the multiple-pitch task could be improved when listening for the lowest or highest voice. Previous studies have shown that listeners are behaviorally most sensitive to pitch changes in the high voice (Palmer &

Holleran, 1994), and even that early brain responses to unexpected stimuli are stronger if the unexpected stimulus comes in the high voice (Fujioka, Trainor, & Ross, 2008). However, some of these previously observed effects might be down to spectral differences, rather than pitch differences, between the voices. In Experiments 5.1-5.3, all stimuli are filtered into the same spectral regions in order to control for spectral overlap between the tones, a measure not usually taken in previous studies. It is also unclear from most previous studies whether the effect is due to the relative or absolute height of the high voice: is the advantage for higher F0s in absolute terms, or for F0s higher than their concurrent neighbors? In Experiment 5.3 we controlled for this by holding target F0 constant and varying the F0s of the maskers.

Beyond the problem of listening to the middle voice, the increased difficulty of the multiple-pitch task in Experiment 5.2 relative to the single-pitch task may be due in part to the increased demand on attention-related processes involved in knowing which tone out of the three to listen for. In an attempt to mitigate this, we used a paradigm that encourages listeners to hear out one tone in the presence of others by repeating the target tone beforehand, as previous studies have done (Darwin, Hukin, & Al-Khatib, 1995; Shinn-Cunningham, Lee, & Oxenham, 2007).

Method

Listeners. Fifteen normal-hearing listeners participated in the experiment, 10 female and five male, ranging in age from 18 to 64 (mean = 30.70), and in years of musical experience from 2 to 13 (mean = 6). All 15 out of 15 listeners passed the pure-tone screening task and completed the other tasks.

Stimuli. The right side of Fig. 5.3 illustrates the paradigm in Experiment 5.3, which was generally similar to Experiment 5.2. Tones were 250 ms in duration with 50 ms gaps in between. The reference tone was presented three times before the target tone, which was presented either alone, or flanked by maskers. The nominal F0 of the reference and target was roved between 262 and 294 Hz, and the difference in F0 between any adjacent pair of tones in the three-tone combination was roved between 3 and 6 semitones, meaning that masker F0s ranged from 131 Hz (12 semitones below minimum nominal F0) to 586 Hz (12 semitones above maximum nominal F0). Importantly, the rove range for reference and target F0s stayed constant and did not change as a function of the target position. Instead, the "low target" condition used a mixture of complexes with a higher average F0 and the "high target" condition used a mixture of complexes with a lower average F0.

Procedure. In Experiment 5.3, instead of the four spectral regions from the other two experiments, tones were presented either as complexes bandpass filtered at 0.5-2 kHz, complexes bandpass filtered at 2-8 kHz, or a full version of the multiple-pitch task with pure tones. Listeners completed 100 trials total per tone type for each target position (and the single pitch task). Tone type changed every block of 20 trials (with block order randomized). Listeners completed all 300 trials for each target position (or single pitch task) before moving on to the next target position, and were explicitly informed in advance of which target position to listen for. The single pitch task was always completed before the multiple pitch tasks, but the order of target positions for the multiple pitch tasks was counterbalanced.

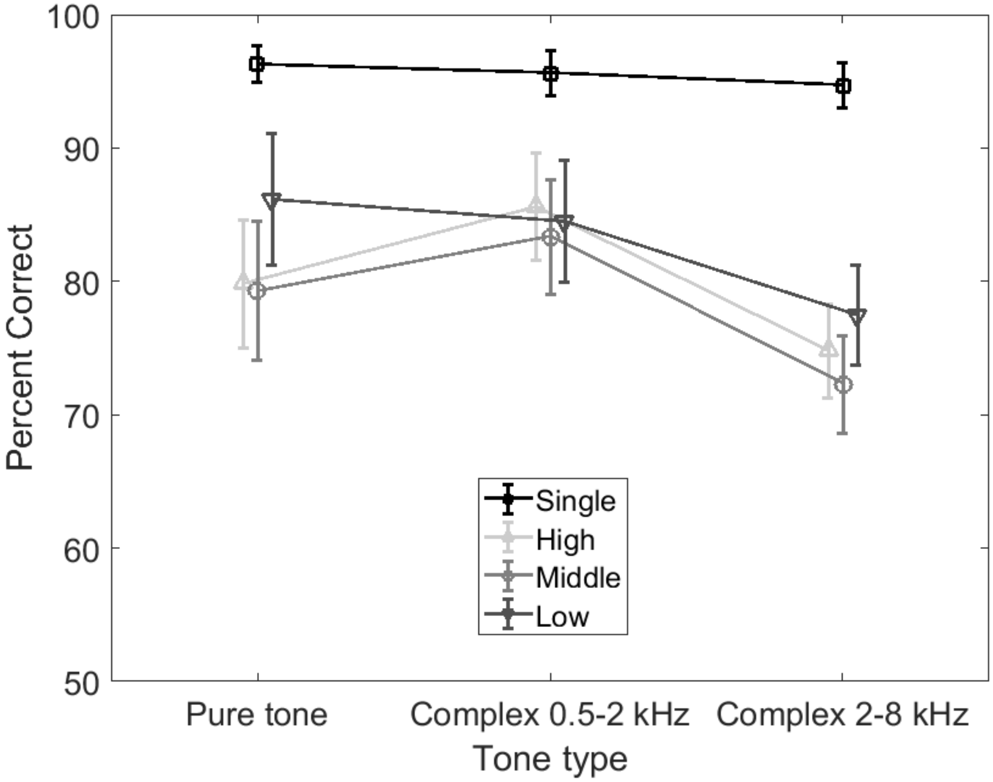


Figure 5.5. Results for Experiment 5.3. Listeners identified the direction of a 0.5-semitone pitch change between reference and target tones, for pure tones or complex tones in one of two spectral regions, with the target tone presented alone, or as the lowest, middle, or highest tone in a three-tone mixture. Error bars show 1 SEM.

Results

The results from Experiment 5.3 are shown in Figure 5.5. Accuracy was near ceiling for all tone types for the single-pitch condition, which was therefore excluded from statistical analysis. A 3 x 3 repeated-measures ANOVA on percent correct scores, transformed into RAU, in the multiple-pitch condition, considering tone type (pure tones, complexes at 0.5-2 kHz, and complexes at 2-8 kHz) and target location (high, middle, and low), found a main effect of tone type [F(2,28) = 20.9, p < 0.001], a main effect of target position [F(2,28) = 4.9, p = 0.02], and an interaction [F(4, 56) = 4.06, p = 0.006]. Post-hoc pairwise comparisons revealed that performance with the low-F0 target trended higher than for the middle-F0 target (p = 0.03), but this trend is insignificant with Bonferroni correction for multiple comparisons. Listeners performed significantly poorer with complexes in the 2-8 kHz spectral region than with complexes in the 0.5-2 kHz regions or with pure tones (p < 0.005 in both cases). Average performance was significantly above chance for all combinations of tone type and target position (p < 0.001 in all cases).

Discussion

Complexes in the 2-8 kHz region, where the combination of tones is unresolved, were discriminated more poorly than complexes in the 0.5-2 kHz region, where the combination is resolved. This could be due to increased harmonic numbers, or to decreased resolvability. However, the fact that a 0.5-semitone pitch change was accurately discriminated even in the 2-8 kHz region suggests that resolvability is not strictly necessary for small pitch DLs.

We did not observe a strong effect of high-voice superiority, possibly suggesting that earlier effects (Fujioka et al., 2008; Palmer & Holleran, 1994) may have been due to spectral differences rather than relative F0 differences between the voices.

Chapter 6: Simultaneous pitches: predictions from models of the auditory system

I. Rate-place theory: Harmonic template matching

We wanted to determine whether the amount of information available to the auditory system through rate-place or temporal codes was sufficient to explain human performance in our behavioral experiments. The initial inputs for both of these methods was simulated neural firing at the level of the auditory nerve (AN) in response to our stimuli, from a model of the auditory periphery (Zilany, Bruce, & Carney, 2014). The model was implemented using code from the UR EAR Modeling Tool (Carney et al., 2016), in MATLAB (The Mathworks, Natick, MA). We modeled 200 center frequencies (CFs) logarithmically spaced between 0.5 and 8 kHz, using human tuning estimates from Shera, Guinan, & Oxenham (2002), modeling AN fibers with medium spontaneous firing rates.

Stimuli and harmonic templates

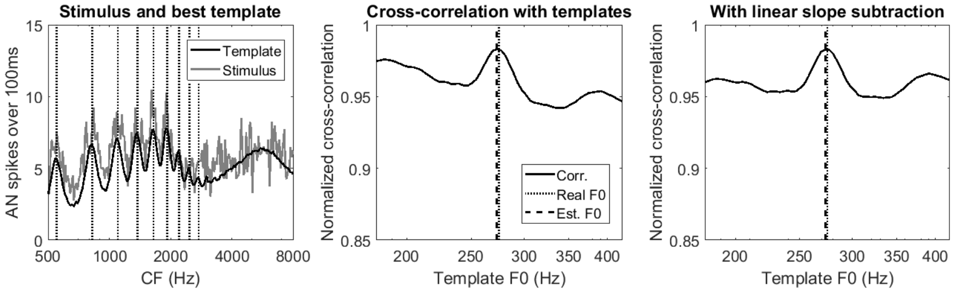


Figure 6.1. Example trial with a single harmonic complex tone using template-matching model. Left: model response to a single harmonic complex tone in the lowest spectral region (0.5-2kHz), and the best-fitting template in that same spectral region. Vertical dashed lines show frequencies of the 2nd-10th components in the presented stimulus. Center: strength of cross-correlations between the stimulus and 60 templates at different F0s. The estimated F0 corresponded to the template with the highest cross-correlation. Right: strength of cross-correlations adjusted to correct for a linear trend.

Model responses were computed for 800 stimuli: 100 harmonic complex tones in each of four spectral regions, for both single tones and triads, with components added in random phase, analogous to the stimuli heard by human listeners in Experiments 5.1-5.3 (see Figure 5.1). For single tones, F0 was sampled on each trial from a uniform distribution between 262 and 294 Hz. For triads, the middle F0 was sampled between 262 and 294 Hz, the upper F0 from a range 3-6 ST above the middle F0 (minimum = 312 Hz, maximum = 416 Hz), and the lower F0 from a range 3-6 ST below the middle F0 (minimum = 185 Hz, maximum = 247 Hz). For modeling purposes, these stimuli were generated with durations of 100 ms, on and off ramps of 20 ms, and a sampling rate of 100 kHz. As in Experiments 5.1-5.3, all stimuli were mixed with TEN, at 10 dB per ERB below the level per component of the stimulus, before being presented to the model. The rate-place representation was obtained by summing the number of spikes across the duration of the stimulus at each CF, then averaging the outputs of 4 AN fibers per CF. This amount of neural noise was chosen by a fitting process, see subsection "Fitting the model to previous data on pitch with resolved and unresolved harmonics." Example model responses to single stimuli in the low spectral region are plotted as the gray lines in Figure 5.1 for a single tone (left) and Figure 5.2 for a triad (left, all three panels).

Instead of looking for salient peaks in the rate-place code representation (as in Micheyl, Keebler, & Oxenham, 2010), we estimated the degree of useful information in this representation by cross-correlating it with harmonic templates, in the style of Larsen, Cedolin, & Delgutte (2008). In each of the four spectral regions, we also computed ideal harmonic templates for 60 F0s, logarithmically spaced between 180 and 420 Hz. Each template was the average of 50 AN model responses with the model in deterministic mode, giving probability of firing rather than spike counts for a given number of fibers. Example templates in the low spectral region are plotted as the black lines in Figure 6.1 (left) and Figure 6.2 (left, all three panels). It should be noted that our method differs from that of Larsen et al. (2008) in that we did not compute harmonic templates for specific combinations of multiple F0s. Instead, we correlated the response to a triad containing multiple F0s with harmonic templates for single F0s. This method seems more plausible in terms of neural implementation, given that it requires fewer stored templates to compare against stimuli.

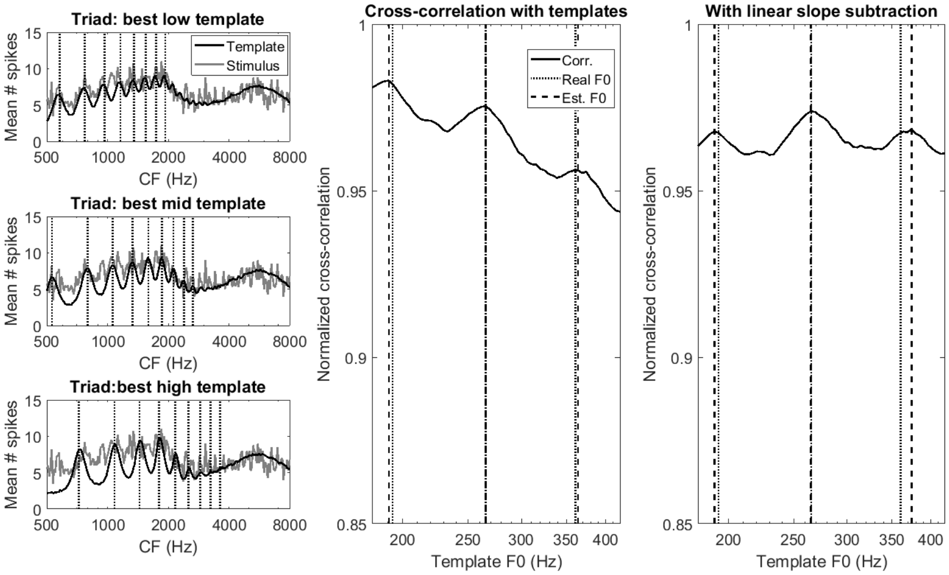


Figure 6.2. Example trial with three concurrent harmonic complex tones using template-matching model. Left: model response to a triad composed of three harmonic complex tones with three different F0s, in the lowest spectral region (0.5-2 kHz), along with the three best-fitting single-pitch templates: low (left, top), middle (left, center), and high (left, bottom). Vertical dashed lines show frequencies of the 2nd-10th components in each of the presented complexes. Center: strength of cross-correlations between the combined stimulus and 60 single-pitch templates at different F0s. The three estimated F0s corresponded to the templates with the three highest peaks of cross-correlation, at least 1 semitone apart. Right: strength of cross-correlations adjusted to correct for a linear trend.

Template matching and F0 estimation

For each of the 800 stimuli, we computed the normalized cross-correlation of the rate-place representation with each of the 60 harmonic templates in the relevant spectral region, in order to create a normalized cross-correlation function (NCCF). The strength of this cross-correlation was taken as an indicator of the strength of the corresponding pitch. For single tones, F0 was estimated to be the maximum of the NCCF. For triads, three F0s were estimated, at the three highest peaks in the NCCF, with the constraint that peaks must be at least 1 ST apart.

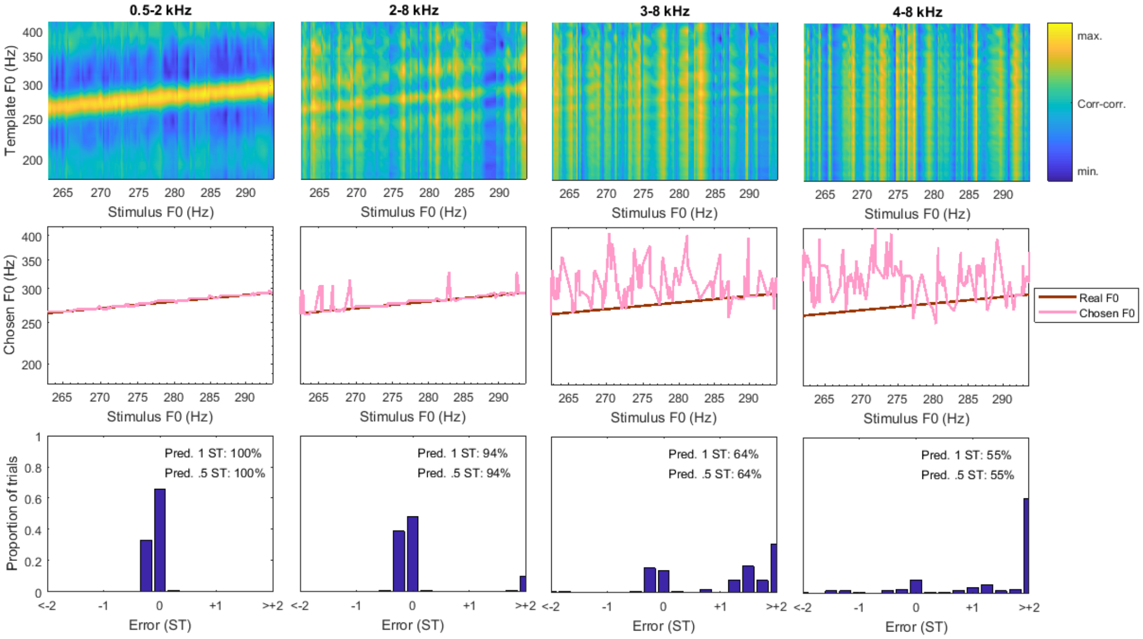


Figure 6.3. Template-matching model outputs for 100 single harmonic complex tones in each of the four spectral regions. Top row: strength of normalized cross-correlations between 100 stimuli and 60 templates. Stimuli were generated by sampling from a uniform distribution of F0s between 262 and 294 Hz, and are sorted here from low to high F0. Templates were spaced logarithmically between 180 and 420 Hz. Middle row: chosen F0 (corresponding to the template with the highest cross-correlation) for each stimulus, along with actual F0 of each stimulus. Bottom row: error distribution of chosen F0s, along with resulting predictions for behavioral performance in a 1-ST and 0.5-ST pitch discrimination task. Predictions are a function of the number of responses within half the distance to be discriminated, and the number of responses predicted to fall in this range by chance.

All NCCFs, for both single tones and triads, exhibited a linear trend towards stronger correlations with templates at lower F0s, possibly due to the shallower modulation depth in low-F0 templates, and hence a larger DC component contributing to the cross-correlation. In order to correct for this trend, we fit a line to each NCCF, then subtracted this line and added the original mean, creating a version of the NCCF with the linear slope removed (see Figures 6.1 and 6.2, right panels). Responses based only on these corrected NCCFs are plotted in Figures 6.3 and 6.4, and behavioral predictions from both corrected and uncorrected versions of the model are shown in Figure 6.6.

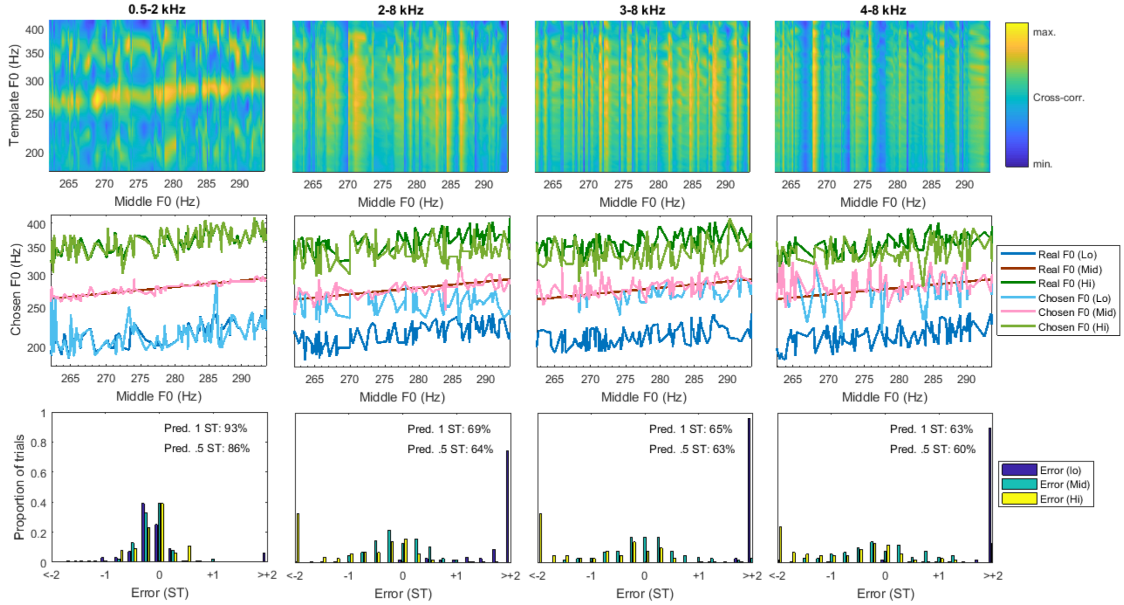


Figure 6.4. Template-matching model outputs for 100 triads, each composed of three harmonic complex tones, in each of the four spectral regions. Top row: strength of normalized cross-correlations between 100 stimuli and 60 templates. Stimuli were generated by sampling middle F0s from a uniform distribution of F0s between 262 and 294 Hz, then sampling high and low F0s from distributions 3-6 ST above and below the middle F0, respectively. Stimuli are sorted here by middle F0, from low to high. Middle row: chosen F0s (corresponding to the three templates with cross-correlation peaks at least 1 ST apart) for each stimulus, along with actual F0s of each stimulus. Bottom row: error distribution for low, middle, and high F0s, along with resulting predictions for behavioral performance in a 1-ST and 0.5-ST pitch discrimination task. Predictions are a function of the mean number of responses (across all three pitches) within half the distance to be discriminated, and the number of responses predicted to fall in this range by chance.

On some trials in response to triads, the model detected a spurious F0 either above the highest actual F0 or below the lowest actual F0. In these cases, the other two estimated F0s sometimes corresponded well to actual F0s at neighboring positions in the triad, e.g. the middle estimated F0 sometimes corresponded well to the highest actual F0. To account for this, the error associated with each estimate was defined as the minimum distance to any actual F0 in the triad, regardless of whether or not the position (low, middle, or high) of the closest actual F0 corresponded to the position of the estimated F0 relative to other estimated F0s.

Although we assumed the error distributions of estimated F0s would be approximately normal, it is worth noting that they appear to be multimodal even for single tones. This may be due to a phenomenon whereby the peaks in a template line up with the peaks in the response to harmonic components shifted one harmonic number away (e.g., the 7th, 8th, and 9th component peaks in a template lining up with the 6th, 7th, and 8th component peaks in the stimulus, respectively). This is not a common type of error in pitch-matching distributions from humans (Oxenham et al., 2011), even in frequency regions beyond the limit of phase locking where access to temporal cues should be severely limited.

Fitting the model to previous data on pitch with resolved and unresolved harmonics

In order to fit the amount of neural noise included in the model, we varied the number of AN fibers per CF from 1 to 20. Spike times were generated independently for each fiber, so the neural representation at each CF became less noisy as spikes were summed across multiple fibers. To choose an appropriate level of neural noise, we selected the fiber count that provided the best fit to data from Houtsma & Smurzynski (1990) on pitch perception of resolved and unresolved harmonics. Based on Exp. III of that paper, we generated missing-fundamental harmonic complex tones with 11 successive components, added in sine phase, with F0 = 200 Hz. We generated 100 stimuli in each of four conditions, where the average lowest harmonic number (N) was the 7th, 10th, 13th, or 19th harmonic. The actual N for each stimulus was roved within +/- 1 from the average for that condition, such that in the condition where average N = 10 (e.g.), roughly 33 out of 100 stimuli included the 9th-19th harmonics, roughly 33 included the 10th-20th, and roughly 33 included the 11th-21st. For our modeling purposes, all stimuli had durations of 100 ms, with 20 ms on- and off-ramps, and were generated with a sampling rate of 100 kHz. Stimuli were presented to the model in the absence of any masking noise.

Figure 6.5A shows the model responses to examples of these stimuli in each condition, for three of the 20 fiber counts. In order to obtain a rate-place representation of the modeled AN response to each stimulus, we summed the number of spikes at each CF across the duration of the stimulus. We then estimated F0 by cross-correlating these responses with 20 harmonic templates, considering only CFs in the spectral region where components were presented. Harmonic templates for these stimuli were the expected response to a broadband harmonic complex including all harmonics under half the sampling rate, obtained by averaging AN firing probability across CFs in response to 50 of these stimuli per template.

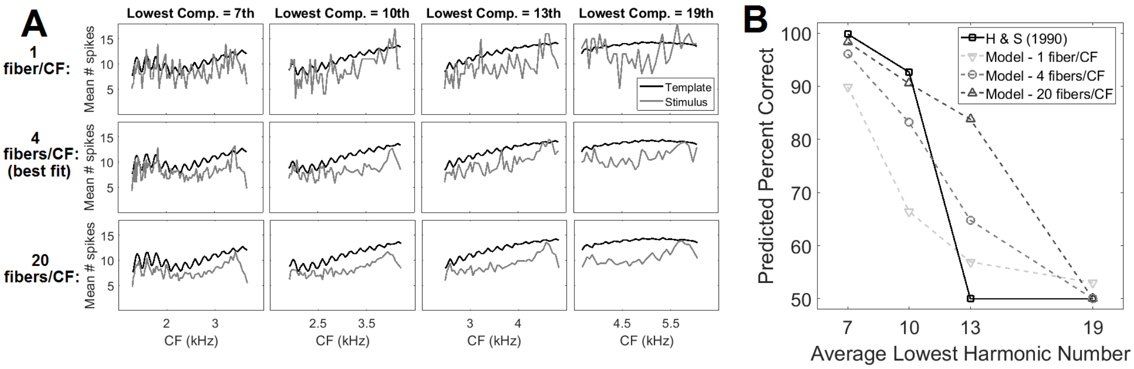


Figure 6.5. Model fitting of neural noise based on data from Houtsma & Smurzynski (1990). (A) Example responses of the Zilany, Bruce, and Carney (2014) AN model to stimuli and relevant templates for harmonic complexes with lowest components at the 7th, 10th, 13th, or 19th harmonic, for three different fiber counts (amounts of neural noise). 100 stimuli were generated in each condition and cross-correlated with 20 templates at different F0s. The template with the highest cross-correlation was selected as the estimated F0, and the error distribution around the correct F0 of 200 Hz was used to compute a predicted percent correct on a 0.5-ST pitch discrimination task. (B) Model predictions for pitch discrimination at 0.5 ST, at three levels of neural noise, compared to predictions based on Houtsma & Smurzynski (1990). Thresholds from that study in the 7th and 10th lowest-component conditions were converted to predicted percent correct on a 0.5-ST discrimination task, as described in subsection A. It was assumed that no rate-place information was available in the 13th or 19th lowest-component conditions. Fiber counts from 1 to 20 were evaluated. The best-fitting amount of neural noise was with 4 fibers/CF, and all subsequent model responses use this parameter.

Based on the results of Exp. III in Houtsma & Smurzynski (1990), F0DLs in the first two conditions (N = 7 and N =10) were roughly 1 Hz and 2 Hz respectively. This gives 0.5% and 1% change in F0 at F0 = 200 Hz, or .09 and .17 ST. To convert these DLs into predicted percent correct on a pitch discrimination task with a constant F0 distance of 0.5 ST, we estimated sensitivity index *d'* as the F0 distance (0.5 ST) divided by the DL, giving estimated *d'* of 5.79 when N = 7 and 2.90 when N = 10. Then we computed percent correct assuming no bias (hit rate = correct rejection rate) for these *d'* values, giving predictions of 99.81% correct when N = 7 and 92.66% correct when N = 10. For the other two conditions, where average N = 13 or 19, there is reason to believe the threshold measured by Houtsma & Smurzynski purely reflects temporal processing, and that at this point there is no useful information in the rate-place code, therefore we assumed this rate-place model should achieve only chance performance (50% correct) when N = 13 or N = 19.

Fig. 6.5B shows how the amount of neural noise in the model was fit to data from Houtsma & Smurzynski (1990). For each fiber count, behavioral predictions were obtained based on the error distribution of estimated F0s around the correct value of 200 Hz (see section D: "Behavioral predictions"). For each fiber count, squared differences between model predictions and data were summed. The fiber count with the lowest sum of squared differences was 4 fibers per CF, so this amount of neural noise was used for all subsequent model responses.

Behavioral predictions

We wanted to compare predictions from this model against behavioral data from Experiment 5.2, where listeners discriminated the direction of a 0.5 ST or 1 ST change in F0, for single tones and triads. In order to convert the error distributions of model F0 estimates into predictions for behavioral performance, we counted the number of responses within a criterion distance, equal to half the distance to be discriminated (i.e. 0.25 ST or 0.5 ST). We then determined the level of chance performance based on a uniform distribution across all 60 templates (3.41% for the 0.25-ST criterion and 6.82% for the 0.5-ST criterion). We defined the chance-corrected proportion within the criterion P_C as equal to (P - C) / (1 - C), where P is the original proportion within the criterion and C is the level of chance performance. We then estimated that on P_c proportion of trials (when the error was within half the distance to be discriminated), accuracy would be 100%, and on the rest of the trials, accuracy would be 50%. Figure 6.6 shows the resulting predictions, for versions of the model with and without linear slope subtraction from NCCFs, compared against behavioral results from Experiment 5.2.

We had fit the amount of neural noise in the model to previous data on single-complex pitch discrimination (Houtsma & Smurzynski, 1990), under the assumption that the model should perform poorly for complexes above the limit of resolvability (N=13 in the Houtsma & Smurzynski study). Correspondingly, the template-matching model significantly under-predicted human performance for single complex tones in the upper two spectral regions. This suggests that the rate-place information needed to account for performance with single harmonics complexes is not sufficient to account for performance with multiple complexes, at least within the framework of the model tested here. There is also a noticeable difference between model and data in the shape of the function for multiple pitch performance: while human performance degraded gradually across the 4 spectral regions, predicted performance from the model declines sharply from the 0.5-2 kHz region to the 2-8 kHz region. This is consistent with the notion that peripheral resolvability for these mixtures is poor, despite the low harmonic numbers of each complex.

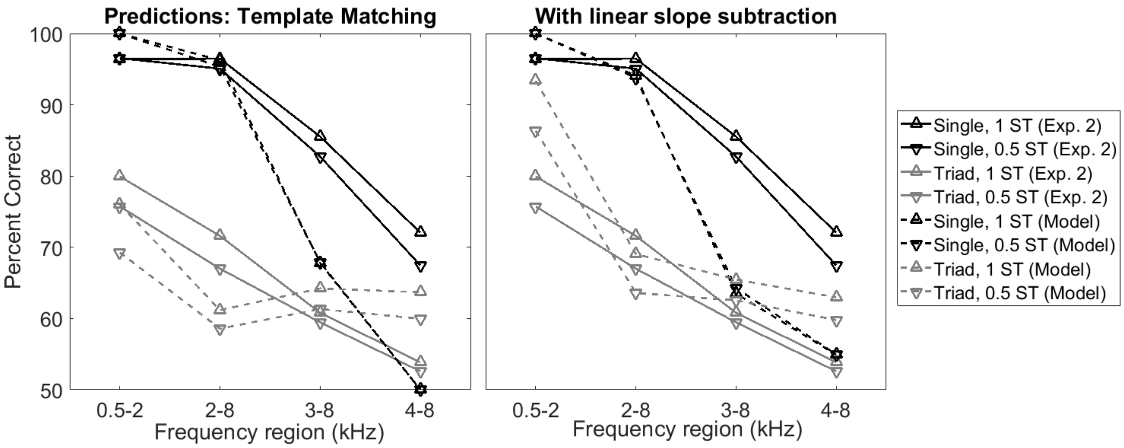


Figure 6.6. Comparison of template-matching model predictions with behavioral results from Exp. 5.2. All predictions obtained using 4 fibers/CF. Performance was predicted to be 100% when the response was within half the distance to be discriminated, and 50% otherwise, after subtracting chance performance. Left: predictions from the basic template-matching model. Right: predictions from the model when a linear slope was subtracted from each cross-correlation function.

II. Temporal theory: summary autocorrelation

A. Summary autocorrelation and F0 estimation

In order to examine the information available from the temporal code, instead of the rate-place code, for these same stimuli, we implemented a summary autocorrelation strategy, after Meddis & O'Mard (1997). The stimuli presented to this model were the same stimuli presented to the template-matching model in the previous section. We started with the same AN responses over time, but instead of summing across time to get a spike count at each CF, we computed autocorrelations of firing patterns over time at each CF, and then summed across CF.

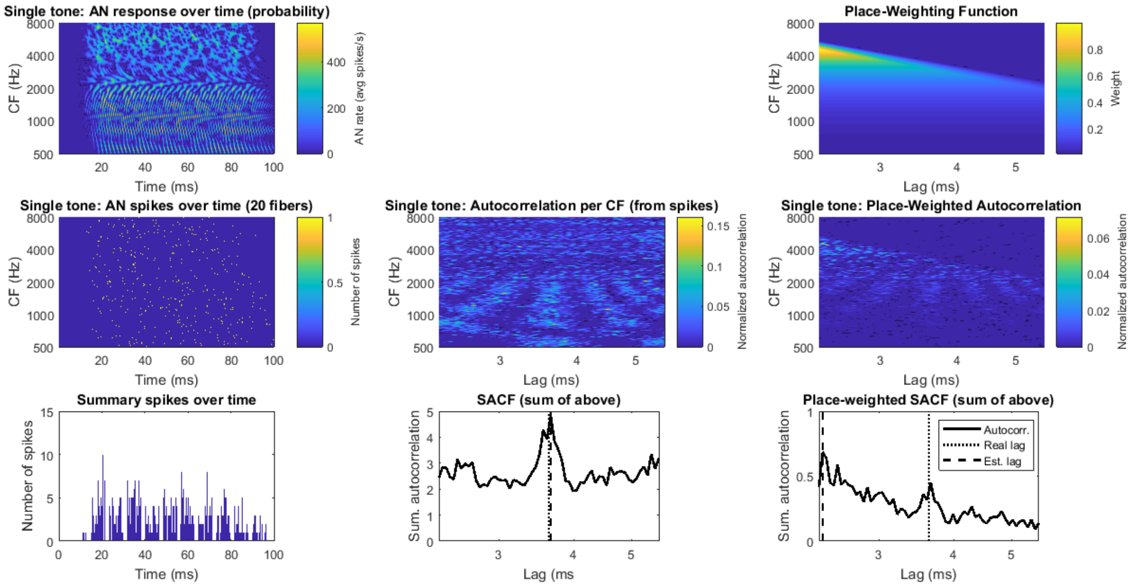


Figure 6.7. Autocorrelation processing of a single harmonic complex tone, in the lowest spectral region (0.5-2 kHz). The AN firing probability over time and across CFs from the Zilany, Bruce, and Carney (2014) model (top left) is shown along with spike times summed across 20 fibers from the same model, separated by CF (center left) and summed across CFs (bottom left). The spike times at each CF are used to compute autocorrelation functions (ACFs, center), which can be weighted using the place-dependent weighting function from Bernstein & Oxenham (2005) (top right) to produce place-weighted ACFs (center right). Summary autocorrelation functions (SACFs) are computed by summing ACFs (bottom center), or by summing place-weighted ACFs (bottom right). Estimated F0 corresponds to the lag at the maximum of the ACF.

The firing pattern over time at each CF was the sum of spikes across 20 fibers (the amount of neural noise chosen by the fitting process, see subsection "Fitting the model to previous data on unresolved pitch"). Autocorrelation functions (ACFs) were computed using 60 lags, corresponding to the inverse of the 60 F0s used for templates in the template-matching model (F0s logarithmically spaced between 180 and 420 Hz). ACFs were computed as in Meddis & O'Mard (1997), with one change: the time point at which autocorrelation was computed was 20 ms before the end of the stimulus, i.e. just before the off ramp. The time constant was 10ms, as in Meddis & O'Mard (1997). ACFs were then summed across CF to create SACFs, examples of which are shown in Figure 6.7 for a single tone (bottom center) and Figure 6.8 for a triad (bottom center).

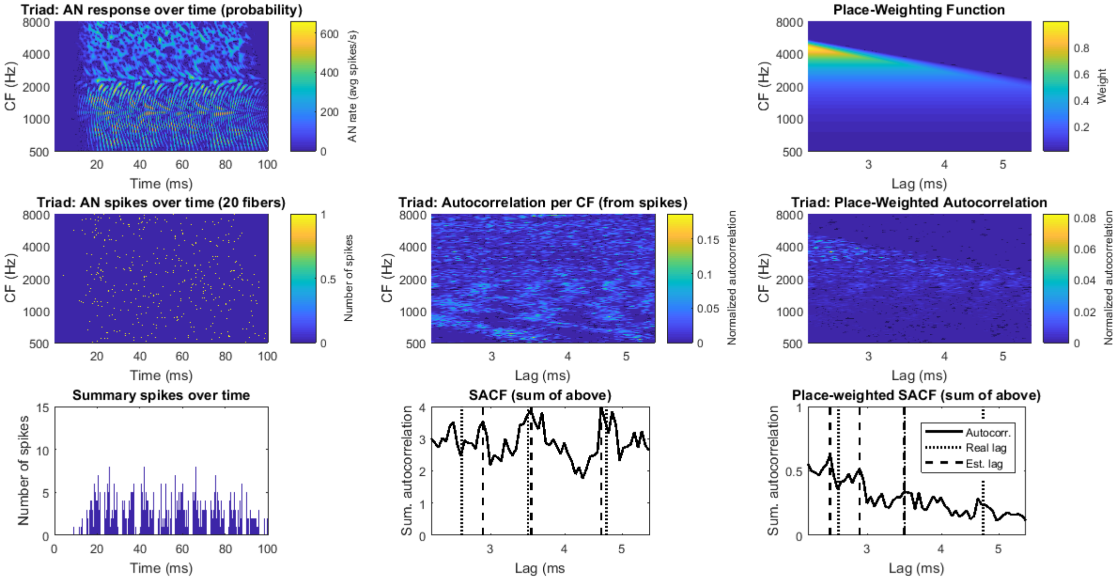


Figure 6.8. Autocorrelation processing of a triad, containing three complex tones with three different F0s, in the lowest spectral region (0.5-2 kHz). Computation of SACFs as in Figure 13. Three F0s were estimated, corresponding to three lags at peaks in the SACFs separated by a difference corresponding to at least a 1-semitone difference in F0.

Given that behavioral data from Experiments 5.1-5.3 did not seem to be strictly limited by peripheral resolvability, we wanted to test whether a place-dependent autocorrelation model, dependent on harmonic number rather than resolvability, would better predict behavioral results. We used the place-weighting function defined by Bernstein & Oxenham (2005) to compute place-weighted SACFs as well as non-place-weighted SACFs. The goal of this procedure is to decrease model performance when CF is far from F0, by limiting the lag window considered at each CF. However, since introduction of place weighting degraded performance severely even for single tones, model responses based only on the non-place-weighted SACFs are plotted in Figures 6.9 and 6.10 Place-weighted predictions are also shown for comparison in Figure 6.12.

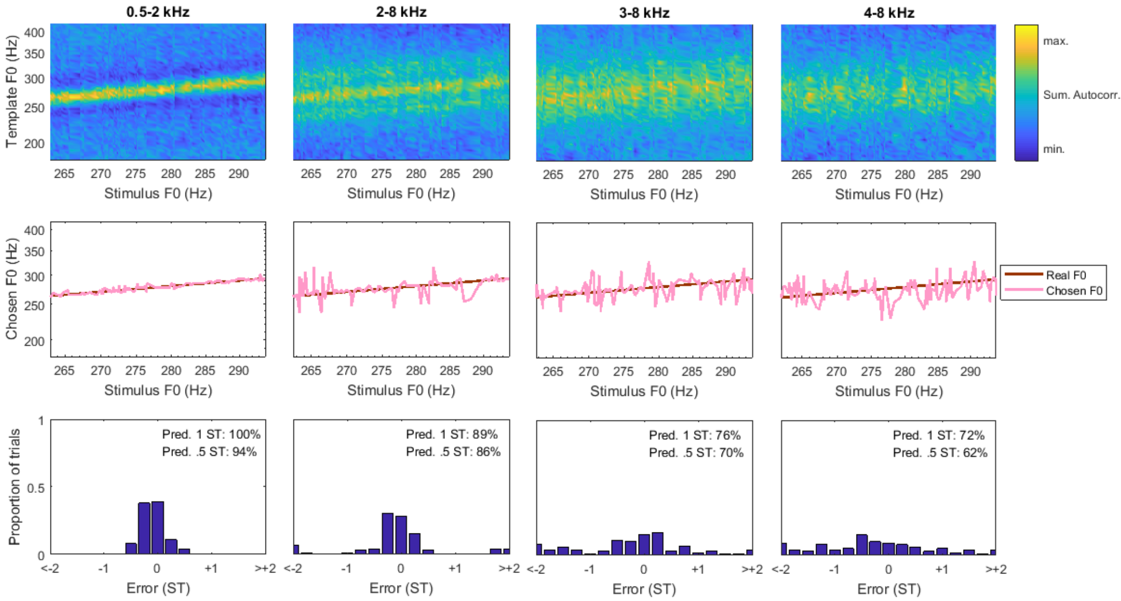


Figure 6.9. Autocorrelation model outputs for 100 single harmonic complex tones in each of the four spectral regions. Top row: summary autocorrelation functions for 100 stimuli, sorted here from low to high F0. Middle row: chosen F0 (the inverse of the lag at the maximum of the SACF) for each stimulus, along with actual F0 of each stimulus. Bottom row: error distribution of chosen F0s, along with resulting predictions for behavioral performance in a 1-ST and 0.5-ST pitch discrimination task. Predictions are computed from error distributions as in the template matching model.

F0 estimates were obtained in the same way as in the template-matching model, using SACFs instead of NCCFs. For single tones, the estimated F0 corresponded to the inverse of the lag at the maximum of the SACF. For triads, the three estimated F0s were the inverses of the three lags at SACF peaks such that chosen F0s were separated by at least 1 semitone. As in the template-matching model, the error associated with each estimate was defined as the minimum distance to any actual F0 in the triad, regardless of position. It is notable that the multi-modality of error distributions in the template-matching model is not apparent in the autocorrelation model, which seems to have more normal distributions.

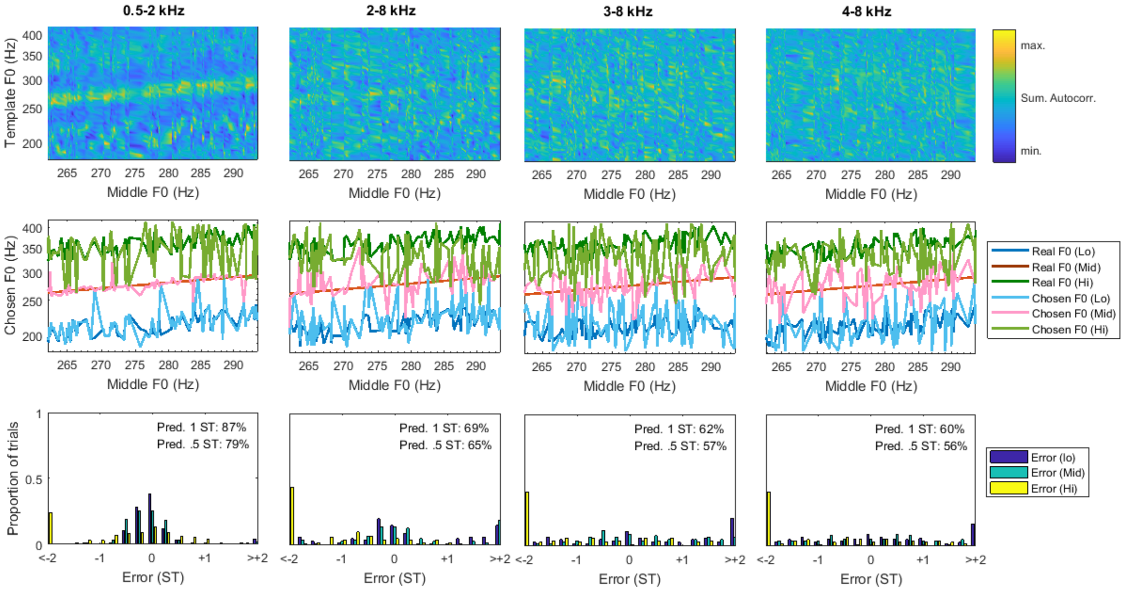


Figure 6.10. Autocorrelation model outputs for 100 triads in each of the four spectral regions. Top row: summary autocorrelation functions for 100 stimuli. Stimuli were generated by sampling middle F0s from a uniform distribution of F0s between 262 and 294 Hz, then sampling high and low F0s from distributions 3-6 ST above and below the middle F0, respectively. Stimuli are sorted here by middle F0, from low to high. Middle row: chosen F0s (corresponding to the inverse of the lags at three peaks in the SACF, at least 1 semitone apart) and actual F0s. Bottom row: error distribution for low, middle, and high F0s, along with resulting predictions for behavioral performance in a 1-ST and 0.5-ST pitch discrimination task.

Fitting the model to previous data on unresolved pitch

We fit the amount of neural noise included in this model separately, to the data from upper conditions in Houtsma & Smurzynski (1990) Exp. III, where N = 13 or N = 19, that was ascribed to temporal mechanisms in the previous section. We generated missing-fundamental harmonic complex tones with 11 successive component with F0 = 200 Hz, added either in sine phase or Schroeder phase, a phase relation that minimizes the peakiness of the temporal envelope, making it minimally useful to the temporal code (in contrast to sine phase, which is maximally peaky). We generated 100 stimuli in each of two conditions, N = 13 or N = 19, for both sine and Schroeder phase relations.

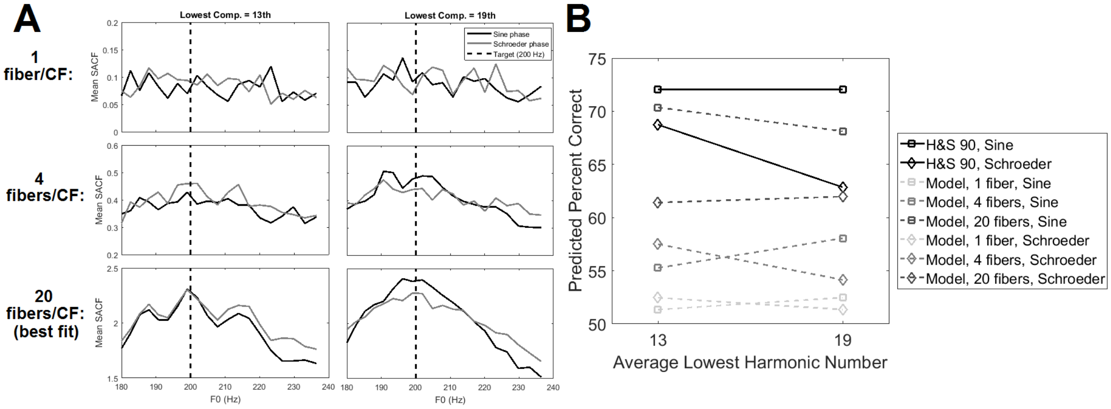


Figure 6.11. Autocorrelation model fitting of neural noise based on data from Houtsma & Smurzynski (1990). (A) Mean SACFs from responses of the Zilany, Bruce, and Carney (2014) AN model to harmonic complexes with lowest components at the 13th or 19th harmonic, for three different fiber counts (amounts of neural noise). On each of 100 trials in each condition, F0 was estimated as the maximum of the SACF, and the error distribution around the correct F0 of 200 Hz was used to compute a predicted percent correct on a 0.5-ST pitch discrimination task. (B) Model predictions for pitch discrimination at 0.5 ST, at three levels of neural noise, compared to predictions based on Houtsma & Smurzynski (1990). Thresholds from that study were converted to predicted percent correct on a 0.5-ST discrimination task, as described in subsection A. Fiber counts from 1 to 20 were evaluated. The best-fitting amount of neural noise was with 20 fibers/CF, and all subsequent model responses use this parameter.

We again varied the number of AN fibers per CF from 1 to 20. Figure 6.11A shows the model responses to examples of these stimuli in each condition and phase relation, for three of the 20 fiber counts. Using the response at each CF, we computed the autocorrelation at 20 different lags corresponding to the 20 F0s evaluated in the fitting stage of the template-matching model. We then summed these autocorrelation functions to produce a summary autocorrelation function (SACF), the maximum of which was used as an estimate of F0.

Based on the results of Exp. III in Houtsma & Smurzynski (1990), F0DLs in sine phase were roughly 5 Hz for both N = 13 and N = 19, but in Schroeder phase were 6 Hz when N = 13 and 9 Hz when N = 19. This gives 0.43, 0.43, 0.51, and 0.76 in terms of ST. When converted into predicted percent correct as in the fitting stage of the template-matching model, this gives 72%, 72%, 69%, and 63% correct on a 0.5-ST pitch discrimination task. Fig. 6.11B shows how the amount of neural noise in the model was fit to data from Houtsma & Smurzynski (1990). For each fiber count, behavioral predictions were obtained based on the error distribution of estimated F0s around the correct value of 200 Hz (see section C: "Behavioral predictions"). For each fiber count, squared differences between model predictions and data were summed. The fiber count with the lowest sum of squared differences was 20 fibers per CF, so this amount of neural noise was used for all subsequent model responses.

Behavioral predictions

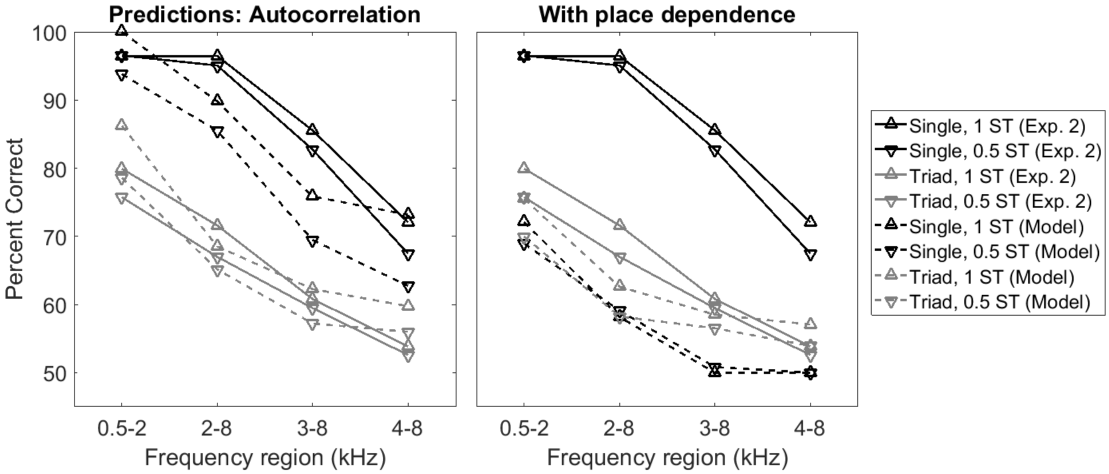


Figure 6.12. Comparison of autocorrelation model predictions with behavioral results from Exp. 5.2. All predictions obtained using 20 fibers/CF. Predictions were calculated from error distributions as with the template matching model. Left: predictions from the basic autocorrelation model. Right: predictions from an autocorrelation model with place dependence, after Bernstein and Oxenham (2005).

As with the template-matching model, we compared predictions from the autocorrelation model against behavioral data from Experiment 5.2. Behavioral predictions were computed from error distributions in exactly the same way as in the template-matching model. Figure 6.12 shows the resulting predictions, for versions of the model with and without a place-dependent weighting function from Bernstein & Oxenham (2005), compared against behavioral results from Experiment 5.2.

The autocorrelation model slightly under-predicted human performance even for single pitches, but it was much more successful at predicting multiple-pitch performance, capturing the more gradual decline observed in human performance, as opposed to the sharp decline between the 0.5-2 kHz and 2-8 kHz regions observed in the template matching model. However, a noticeable difference between the autocorrelation and template-matching models is the large predicted difference between the 1 ST and 0.5 ST tasks, larger than observed in the data from Experiment 5.2. This may have to do with the shape of error distributions, which are normal in the autocorrelation model as opposed to the multi-modal distributions in the template-matching model. Introducing place dependence degraded performance too severely to produce meaningful behavioral predictions to compare with data.

III. Predictions from a combined spectrotemporal model.

Both the spectral (template-matching) and temporal (autocorrelation) models evaluated here failed to capture certain aspects of human performance on their own. The template-matching model under-predicted performance for single pitches in the high, unresolved conditions, and predicted a large difference between the first and second conditions that was not observed in human behavioral data. The autocorrelation model, on the other hand, generally under-predicted human performance for single pitch discrimination.

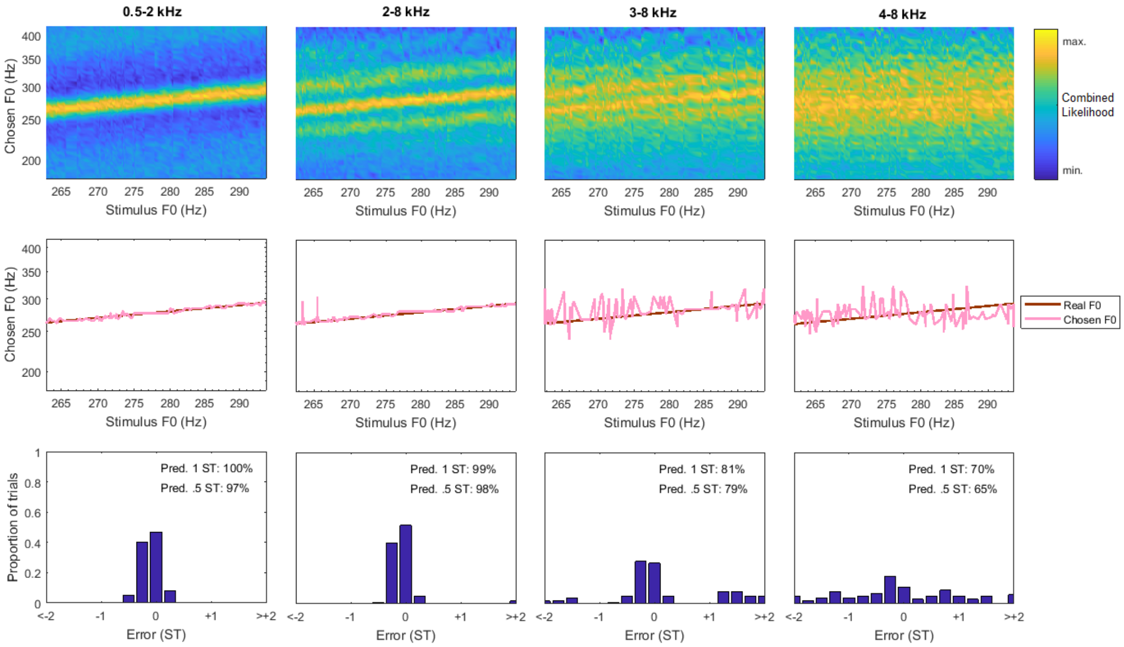


Figure 6.13. Combined spectrotemporal model outputs for 100 single harmonic complex tones in each of the four spectral regions. Top row: combined likelihood functions for 100 stimuli, computed by normalizing and summing NCCFs and SACFs. Middle row: chosen F0 (maximum of combined likelihood function) for each stimulus, along with actual F0 of each stimulus. Bottom row: error distribution of chosen F0s, along with resulting predictions for behavioral performance in a 1-ST and 0.5-ST pitch discrimination task.

In an attempt to arrive at a single model that better describes the behavioral data, we also tested the performance of a model that combines information from the rate-place and temporal codes. This was achieved by normalizing and summing together the F0 likelihood functions of the two models: the NCCFs from the template-matching model and the SACFs from the autocorrelation model, mapped from lag to corresponding F0 (inverse of lag). We used the version of the template-matching model that included linear slope correction, and the version of the autocorrelation model that did not include place-dependent lag weighting. Amounts of neural noise in the two models were different, according to the fitting process for each model to relevant data from Houtsma &

Smurzynski (1990). Specifically, the template-matching model used only 4 fibers/CF, while the autocorrelation model used 20 fibers/CF.

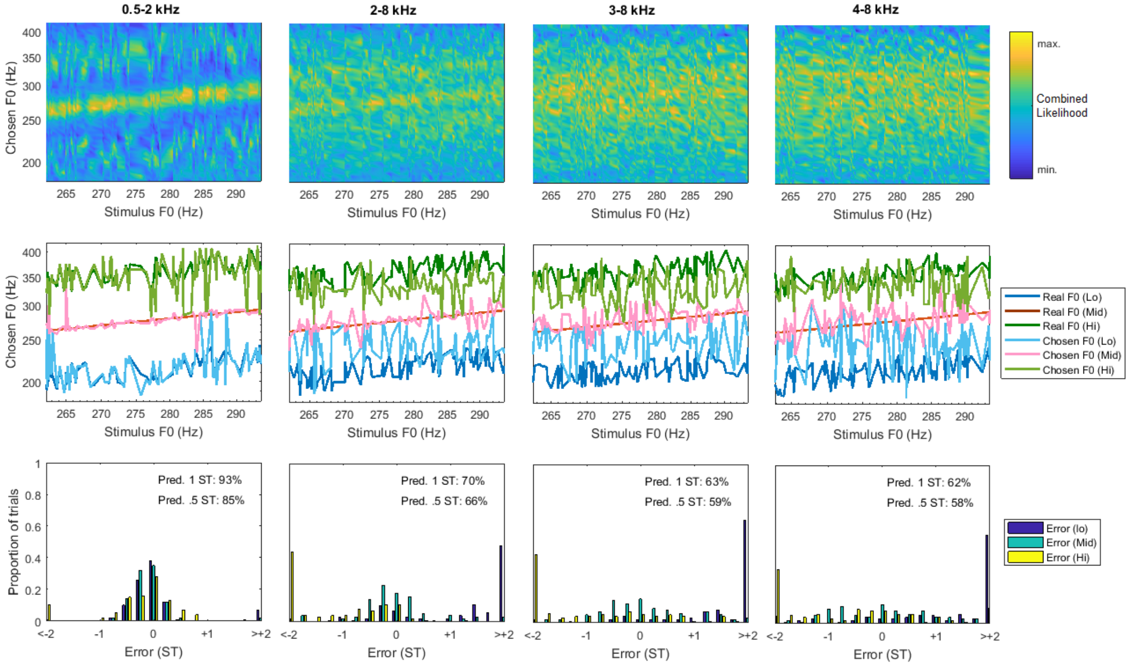


Figure 6.14. Combined spectrotemporal model outputs for 100 triads, each composed of three harmonic complex tones, in each of the four spectral regions. Top row: combined likelihood functions for 100 stimuli per condition. Stimuli are sorted here by middle F0, from low to high. Middle row: chosen F0s (corresponding to three combined likelihood peaks at least 1 ST apart) for each stimulus, along with actual F0s of each stimulus. Bottom row: error distribution for low, middle, and high F0s, along with resulting predictions for behavioral performance in a 1-ST and 0.5-ST pitch discrimination task.

For each of the 100 stimuli in each condition, the SACF and NCCF were both normalized by first subtracting the mean of the function, then dividing by the standard deviation, such that the two normalized functions had means of 0 and standard deviations of 1. They were then added together to produce a combined F0 likelihood function. This combined likelihood function was treated in the same way as the NCCFs in the template-matching model or the SACFs in the autocorrelation model: F0s were chosen based on its peaks, and behavioral predictions were made from error distributions of chosen F0s around real F0s. Figure 6.13 shows model results for single tones, and Figure 6.14 shows results for triads.

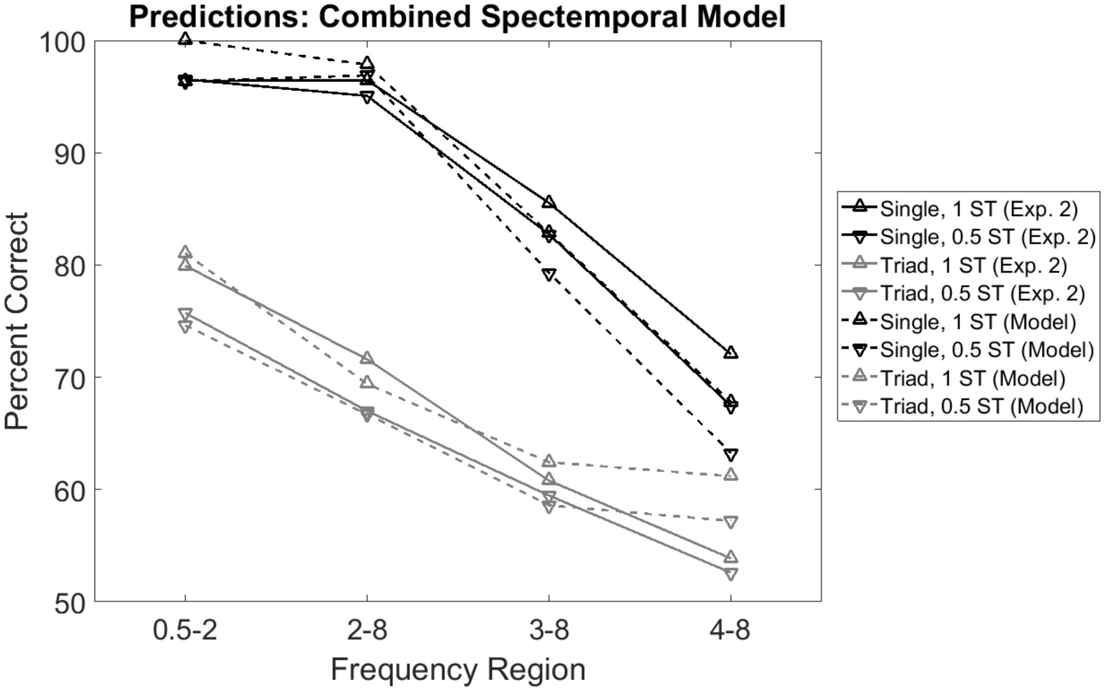


Figure 6.15. Comparison of combined spectrotemporal model predictions with behavioral results from Exp. 5.2. Predictions were calculated from error distributions of F0s chosen using combined likelihood functions from normalizing and summing NCCFs with linear slope correction and SACFs with no place dependence.

Figure 6.15 compares behavioral predictions from this combined spectrotemporal model to actual human performance in Experiment 2. This model provides a better fit to the data than either the template-matching model or the autocorrelation model on their own, providing the minimum sum of squared differences from behavioral data. Computed on percent correct units, the sum of squared differences was 1786 for the template-matching model (with linear slope correction), 564 for the autocorrelation model (without place dependence), and 159 for the combined spectrotemporal model.

IV. Conclusions

The results of our behavioral experiments generally suggest that sub-semitone pitch discrimination is possible even for mixtures of three concurrent complex tones filtered into spectral regions where the mixture contains no resolved harmonics. Briefly, addition of simultaneous spectrally overlapping complex tones does not produce the deterioration in pitch perception one would expect if resolvability was critically important to accurate pitch perception. This has interesting implications for rate-place and temporal models of pitch perception, which have diverging explanations for the phenomenon of sharply degraded pitch perception for missing-F0 complex tones when about the 10th harmonic is removed (Houtsma & Smurzynski, 1990). The standard explanation for this phenomenon, grounded in the rate-place code, holds that the degraded pitch perception is due to loss of resolved harmonics due to broadening auditory filters (Glasberg & Moore, 1990).

Both rate-place and temporal models failed to predict certain aspects of human pitch perception of three-complex mixtures, though they generally predicted the pattern of decreasing performance with decreasing resolvability of harmonic components. This suggests that neither rate-place nor temporal information is sufficient on its own to explain multiple pitch perception. Combining these two kinds of information, however, produced a much better approximation of human behavior. This suggests that multiple concurrent complex pitches may represent a class of stimuli for which both rate-place and temporal information must be accurately encoded to allow for normal human pitch perception.

This study expanded the set of stimuli used in psychoacoustic research to include mixtures of three simultaneous spectrally overlapping harmonic complex tones. It is easy to imagine using even more than three harmonic complexes at once. In music, many more than three distinct pitches are often heard simultaneously. One direction for further exploration of multiple complex pitch perception is to test the limits of this phenomenon. For low-numbered harmonics, how many different pitches can be played simultaneously in an overlapping spectral region before pitch perception begins to deteriorate?

Chapter 7: General summary and conclusions

In everyday life, it is frequently necessary for humans to compare one pitch to its neighbors, occurring either concurrently or in the recent past. This ability is necessary to perceive pitch contours in speech and music, and to perceive musical chords and harmony. The studies in this dissertation explored the processes that underlie perception of pitch relationships, both sequential and simultaneous. The studies investigating sequential pitch relationships provided insight into mechanisms for pitch memory and expectation, while the studies investigating simultaneous pitch relationships furthered our understanding of how pitch may be coded in the auditory system.

In Chapters 2 and 3, we observed evidence that normal-hearing listeners perceive brightness and loudness contours in a manner broadly similar to pitch contours, but that where differences between dimensions exist, they may be less pronounced for listeners with congenital amusia. The conclusion from Chapter 2, that these auditory dimensions may share a mechanism for generating expectations, is slightly surprising given early work showing that sequential processing is generally less accurate for non-speech-related and non-musical auditory stimuli (Warren, Obusek, Farmer, & Warren, 1968). The possibly shared mechanism for expectation could even be broadened even beyond the sense of hearing. Certain similarities have been observed between the perceptual dimensions of loudness (related to sound intensity) and visual brightness (related to light intensity), such as their similar interactions with stimulus duration (Stevens & Hall, 1966). Audiovisual integration for other complex tasks such as speech perception (e.g. McGurk & MacDonald, 1976) suggests that the two modalities may share some mechanisms. It seems plausible that basic expectations for stimulus continuation, such as small intervals and regression to the mean, are domain-general, and would also extend to visual dimensions such as brightness. Future research could examine cross-modal perception of contour using both auditory and visual stimuli.

If pitch, brightness, and loudness do share a mechanism for contour perception and memory, however, it is difficult to explain the finding from Chapter 3 that amusics' short-term memory deficit for pitch is not mirrored in loudness. If the same mechanism is responsible for representing contour in all three dimensions, any deficit in pitch contour perception should have been equally observed in brightness and loudness. Electrophysiological evidence suggests that the pitch deficit in amusia may arise at the level of awareness, with initial detection intact (Moreau, Jolicœur, & Peretz, 2013; Peretz, Brattico, Järvenpää, & Tervaniemi, 2009). If amusics are less impaired in memory for loudness contours, would this translate to a normal P3 response to loudness changes, signaling normal awareness, in contrast to the reduced P3 to pitch changes? Future studies could answer this question with electrophysiological measures.

In some cases, the limits of pitch perception and discrimination can be extended through training or context effects, as observed in Chapter 4. When it comes to long-term training effects, even basic F0DLs can be dramatically reduced with a few hours of practice (Micheyl, Delhommeau, Perrot, & Oxenham, 2006). On a short-term basis, the results of Chapter 4 showed that priming a familiar tonal hierarchy had a similar, if smaller, effect on pitch interval perception. These hierarchies are learned through exposure, but some preference towards unequal organization of pitches within the octave may be innate (Trehub, Schellenberg, & Kamenetsky, 1999).

In general, complex pitch perception seems highly dependent on context. In the special case where the tones to be discriminated are Shepard tones, containing only octave frequencies, the perception of pitch height is very strongly and persistently affected by recent context (Chambers et al., 2017; Pelofi, de Gardelle, Egré, & Pressnitzer, 2017). Even basic aspects of complex pitch perception such as the synthesis of harmonic components and the relative dominance of components (Moore, 1985) can be affected by context (Gockel, Alsindi, Hardy, & Carlyon, 2017). All of this suggests that complex pitch perception is, at the least, susceptible to top-down influences dependent on both long-term learning and short-term context.

Before considering top-down influences, however, a current priority for auditory scientists is to better understand complex pitch perception from the bottom up. After all, even high-level pitch sequence processing is also influenced by peripheral factors such as harmonic resolvability (Cousineau, Demany, & Pressnitzer, 2010). The results of Chapter 5 present a challenge for existing peripheral models of pitch perception, some of which were explored in Chapter 6 in light of these results.

There are at least two different possible representations of pitch in the auditory system, the rate-place code and the time code. The idea that these two representations might function as separate, independent pitch mechanisms has its most basic support from studies of pure tones. The perceptual limit of 5 kHz is suggestive of an upper limit of phase locking: above this limit, FDLs are significantly elevated (Moore, 1973), and it also corresponds with the limit of musical pitch for pure tones (Attneave & Olson, 1971), though complex tones with harmonic components in this spectral region can produce melodic pitch (Oxenham et al., 2011).

The conflict between rate-place and temporal models of pitch perception runs through the research on the phenomenon of sharply degraded pitch perception for missing-F0 complex tones when about the 10th harmonic is removed (Houtsma & Smurzynski, 1990). The standard explanation for this phenomenon is grounded in the rate-place code, holding that the degraded pitch perception is due to loss of resolved harmonics from broadening auditory filters (Glasberg & Moore, 1990). This view is supported by the emergence of phase effects for unresolved pitch, indicating that component waveforms are interacting and summing in the neural response. However, the relationship between peripheral resolvability and F0DLs is not perfect: dichotic presentation improves resolvability without improving DLs (Bernstein & Oxenham, 2003), while mistuning odd harmonics by 3% improves DLs without improving resolvability (Bernstein & Oxenham, 2008). Chapter 5 provided additional evidence for the argument that pitch discrimination is not strictly limited by harmonic resolvability, by more clearly dissociating resolvability from harmonic number with the use of three-complex mixtures. Listeners generally performed well in conditions where single complexes would be resolved, even when listening to a mixture with drastically decreased resolvability.

For single complex tones, it had seemed that adding place dependence into a temporal model of pitch perception (Bernstein & Oxenham, 2005) might allow autocorrelation-based temporal models to explain the weakening of pitch for unresolved harmonics, as well as the finding that incongruent place information disrupts pitch sensation (Oxenham et al., 2004). However, in Chapter 6, we found that for the stimuli used in Chapter 5, degrading the autocorrelation model by limiting CF in this way only made it even less effective, being unable to predict even single-pitch performance. A standard autocorrelation model came closer to predicting human performance, but still under-predicted performance across the board. An additional problem for all autocorrelation models of pitch perception is the difficulty of implementing neural delays long enough to detect low pitches – while these delays are likely not physically present, they may be synthesized with phase interactions (de Cheveigné & Pressnitzer, 2006).

The second model, which also came reasonably close to predicting human performance, was a harmonic template matching model in the style of Cedolin & Delgutte (2005). Single harmonic templates were successfully matched to stimuli, even triads containing multiple pitches. The simple approach of correlation with single templates does not require storage of templates for every possible combination of 2 (or 3, or more) F0s, which had been a disadvantage for models using combination templates (Larsen et al., 2008), as storing all combination templates would be costly. Since it is fundamentally a rate-place code representation, the template-matching model also has the advantage of agreeing with basic psychophysical findings such as the general dominance of low, resolved harmonics for the pitch of a complex tone (Plomp, 1967; Ritsma, 1967).

The best results in Chapter 6 came from a combination of the template matching with the temporal model, providing a better approximation to behavioral data than either model alone. This suggests that it may be necessary to include both spectral and temporal information in order to account for human pitch perception, or at least to account for pitch discrimination in mixtures of three concurrent complex tones.

The resolvability of individual components within a harmonic complex can to some extent be measured behaviorally (Moore, Glasberg, & Shailer, 1984). But focusing only on resolvability misses the fact that components are also normally hidden within complexes by informational masking; they can be made more salient by mistuning (Moore, Glasberg, & Peters, 1986), or by pulsing of a probe tone (Moore, Glasberg, & Oxenham, 2012), in a strategy similar to the one tried in this dissertation in Experiment 5.3 by pulsing a complex probe before a complex mixture. Future studies could examine whether the resolvability of the complex mixtures used in Chapter 5, when measured behaviorally in this manner, agrees with predictions from rate-place models about resolvability. If poor resolvability were observed despite accurate pitch discrimination of the complexes within the mixture, the argument would be more convincing for a spectrotemporal model, or at least a model that goes beyond a temporally frozen rate-place representation.

Pitch perception, like many aspects of perception, can be plastic, as demonstrated recently in cochlear-implant users (Reiss, Turner, Karsten, & Gantz, 2013). If the process of assigning a pitch to peripheral input is plastic, it is easy to imagine that much of pitch perception may be learned early in life through experience. Certain models of pitch perception have been proposed using unsupervised neural networks (e.g. Ahmad, Higgins, Walker, & Stringer, 2016), allowing inputs to be matched to pitch sensations through trial and error. Future work could test whether a neural network of this kind, when trained on single pitches, would be perform well when tested on multiple simultaneous pitches, or whether it would require training on simultaneous pitches in the first place.

Other important tests of any working model of pitch perception would come in the form of modeling effects of hearing loss. A model that explains multiple pitch performance would not only have to account for resolved and unresolved pitch, but also the known effect of sensorineural hearing loss on resolvability (Bernstein & Oxenham, 2006b). It also seems plausible that a sensitive task like multiple pitch detection could be susceptible to subtler forms of hearing loss such as synaptopathy (Kujawa & Liberman, 2009), so it may be worth exploring whether the stimuli used in Chapter 5 are experienced differently by those with and without synaptopathic "hidden" hearing loss.

While many aspects of human pitch perception (e.g. the dominance of relative pitch) are likely to be at least partly specific to humans, primates have been shown to share many basic features of human complex pitch perception (Song, Osmanski, Guo, & Wang, 2015). Does this similarity extend to relationships between complex pitches? Some primates can be trained to discriminate the direction of a pitch change (Brosch, Selezneva, Bucks, & Scheich, 2004), but do they develop expectations for continuation of pitch contours? Can they store pitch contours in memory? And are they capable of multiple pitch perception like the kind demonstrated in Chapter 5? Future research could explore to what extent the abilities of relative pitch perception demonstrated in this dissertation are truly specific to humans, and to what extent models of human pitch perception may also apply to close cousins such as primates.

The results reported in this dissertation suggest that human perception of pitch is sensitive to relationships with recent or concurrent pitches. Perception of pitch contours and generation of contour-based expectations seems generalizable to other auditory dimensions; however, a deficit exists in a sub-population identified as amusic, which does not appear to generalize to loudness. Pitch interval perception can be enhanced by the presence of familiar tonal context, but this effect is small. Spectral and temporal models of pitch perception both fail on their own to fully explain human perception of three concurrent pitches, but a model that combines both types of information had more success. Overall, these findings suggest that pitch perception involves bottom-up integration of both spectral and temporal information, as well as top-down effects of learning and context.

References

Ahmad, N., Higgins, I., Walker, K. M. M., & Stringer, S. M. (2016). Harmonic Training and the Formation of Pitch Representation in a Neural Network Model of the Auditory Brain. *Frontiers in Computational Neuroscience*, *10*(March), 24. http://doi.org/10.3389/fncom.2016.00024

Albouy, P., Cousineau, M., Caclin, A., Tillmann, B., & Peretz, I. (2016). Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia. *Scientific Reports*, *6*(1), 18861. http://doi.org/10.1038/srep18861

Albouy, P., Schulze, K., Caclin, A., & Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain Research*, *1537*, 224–232. http://doi.org/10.1016/j.brainres.2013.09.003

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *The Journal of the Acoustical Society of America*, *135*(3), 1371–9. http://doi.org/10.1121/1.4863269

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology*, *84*(2), 147–166. http://doi.org/10.2307/1421351

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain : A Journal of Neurology*, *125*(Pt 2), 238–51.

Balch, W. (1981). The role of symmetry in the good continuation ratings of two-part tonal melodies. *Perception & Psychophysics*, *29*(I), 47–55.

Beerends, J. G., & Houtsma, a J. (1986). Pitch identification of simultaneous dichotic two-tone complexes. *Journal of the Acoustical Society of America*, *80*(4), 1048–1056. http://doi.org/10.1121/1.397974

Beerends, J. G., & Houtsma, A. J. M. (1989). Pitch identification of simultaneous diotic and dichotic two□tone complexes. *The Journal of the Acoustical Society of America*, *85*(2), 813–9.

Bernstein, J. G. W., & Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *The Journal of the Acoustical Society of America*, *113*(6), 3323–3334. http://doi.org/10.1121/1.1572146

Bernstein, J. G. W., & Oxenham, A. J. (2005). An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. *The Journal of the Acoustical Society of America*, *117*(6), 3816–3831. http://doi.org/10.1121/1.1904268

Bernstein, J. G. W., & Oxenham, A. J. (2006a). The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level. *The Journal of the Acoustical Society of America*, *120*(6), 3916. http://doi.org/10.1121/1.2372451

Bernstein, J. G. W., & Oxenham, A. J. (2006b). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *The Journal of the Acoustical Society of America*, *120*(6), 3929. http://doi.org/10.1121/1.2372452

Bernstein, J. G. W., & Oxenham, A. J. (2008). Harmonic segregation through mistuning can improve fundamental frequency discrimination. *The Journal of the Acoustical Society of America*, *124*(3), 1653–67. http://doi.org/10.1121/1.2956484

Bharucha, J. J. (1987). Music Cognition and Perceptual Faciliation: A Connectionist Framework. *Music Perception*, *5*(1), 1–30.

Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception & Psychophysics*, *59*(7), 1098–107.

Bolles, R. C. (1972). Reinforcement, expectancy, and learning. *Psychological Review*, *79*(5), 394–409. http://doi.org/10.1037/h0033120

Boltz, M. (1991). Some structural determinants of melody recall. *Memory & Cognition*, *19*(3), 239–251. http://doi.org/10.3758/BF03211148

Borchert, E. M. O., Micheyl, C., & Oxenham, A. J. (2011). Perceptual grouping affects pitch judgments across time and frequency. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(1), 257–69. http://doi.org/10.1037/a0020670

Borchert, E., & Oxenham, A. (2010). Musical context affects detection of pitch differences in tones with different spectra. *The Journal of the Acoustical Society of America*, *127*(3), 1989.

Bregman, A. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences*, *113*(6), 1666–1671. http://doi.org/10.1073/pnas.1515380113

Brosch, M., Selezneva, E., Bucks, C., & Scheich, H. (2004). Macaque monkeys discriminate pitch relationships. *Cognition*, *91*(3), 259–272. http://doi.org/10.1016/j.cognition.2003.09.005

Burns, E. M., & Viemeister, N. F. (1976). Nonspectral pitch. *The Journal of the Acoustical Society of America*, *60*, 863–869.

Burns, E. M., & Ward, W. D. (1978). Categorical perception--phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America*, *63*(2), 456–468. http://doi.org/10.1121/1.381737

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology: Music, Mind & Brain*, *1*, 12–29.

Carlyon, R. P. (1996). Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *The Journal of the Acoustical Society of America*, *99*(1), 517–24. http://doi.org/10.1121/1.414510

Carney, L. H., Fan, L., Galant, N., Maxwell, B., Teverovsky, D., & Varner, T. (2016). UR EAR Modeling Tool.

Cedolin, L., & Delgutte, B. (2005). Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *Journal of Neurophysiology*, *94*(1), 347–362. http://doi.org/10.1152/jn.01114.2004

Chambers, C., Akram, S., Adam, V., Pelofi, C., Sahani, M., Shamma, S., & Pressnitzer, D. (2017). Prior context in audition informs binding and shapes simple features. *Nature Communications*, *8*, 15027. http://doi.org/10.1038/ncomms15027

Cohen, A., Thorpe, L., & Trehub, S. (1987). Infants' perception of musical relations in short transposed tone sequences. *Canadian Journal of Psychology*, *41*(1), 33–47.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.).

Hillsdale, NJ: Lawrence Earlbaum Associates.

Congdon, P. (2006). *Bayesian Statistical Modelling*. Chichester, England: Wiley.

Cousineau, M., Demany, L., & Pressnitzer, D. (2009). What makes a melody: The perceptual singularity of pitch sequences. *The Journal of the Acoustical Society of America*, *126*(6), 3179–87. http://doi.org/10.1121/1.3257206

Cousineau, M., Demany, L., & Pressnitzer, D. (2010). The role of peripheral resolvability in pitch-sequence processing. *The Journal of the Acoustical Society of America*, *128*(5), EL236–L241. http://doi.org/10.1121/1.3499701

Cousineau, M., McDermott, J. H., & Peretz, I. (2012). The basis of musical consonance as revealed by congenital amusia. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), 19858–63. http://doi.org/10.1073/pnas.1207989109

Cousineau, M., Oxenham, A. J., & Peretz, I. (2015). Congenital amusia: a cognitive disorder limited to resolved harmonics and with no peripheral basis. *Neuropsychologia*, *66*, 293–301. http://doi.org/10.1002/ana.22528.Toll-like

Cuddy, L. L., & Dobbins, P. A. (1988). Octave discrimination: Temporal and contextual effects. *Canadian Acoustics*, *16*(3), 3–13.

Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals. *Perception & Psychophysics*, *57*(4), 451–462.

D'Amato, M. R. (1988). A search for tonal pattern perception in Cebus monkeys: Why monkeys can't hum a tune. *Music Perception*, *5*(4), 453–480. http://doi.org/10.2307/40285410

Darwin, C. J., Hukin, R. W., & Al-Khatib, B. Y. (1995). Grouping in pitch perception: Evidence for sequential constraints. *Journal of the Acoustical Society of America*, *98*(2 I), 880–885. http://doi.org/10.1121/1.413513

de Cheveigné, A., & Kawahara, H. (1999). Multiple period estimation and pitch perception model. *Speech Communication*, *27*(3), 175–185.

de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, *111*(4), 1917. http://doi.org/10.1121/1.1458024

de Cheveigné, A., & Pressnitzer, D. (2006). The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction. *The Journal of the Acoustical Society of America*, *119*(6), 3908. http://doi.org/10.1121/1.2195291

Demany, L., & Semal, C. (1992). Detection of inharmonicity in dichotic pure-tone dyads. *Hearing Research*, *61*(1-2), 161–6.

Demany, L., Semal, C., & Pressnitzer, D. (2011). Implicit versus explicit frequency comparisons: two mechanisms of auditory change detection. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(2), 597–605. http://doi.org/10.1037/a0020368

Devergie, A., Grimault, N., Tillmann, B., & Berthommier, F. (2010). Effect of rhythmic attention on the segregation of interleaved melodies. *The Journal of the Acoustical Society of America*, *128*(1), EL1–EL7. http://doi.org/10.1121/1.3436498

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341–354.

Dowling, W. J. (1986). Context effects on melody recognition: Scale-step versus interval representations. *Music Perception*, *3*(3), 281–296. http://doi.org/10.2307/40285338

Dowling, W. J. (1990). Expectancy and attention in melody perception. *Psychomusicology: Music, Mind & Brain*, *9*(2), 148–160.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*(2), 524–531.

Dowling, W. J., & Tillmann, B. (2014). Memory Improvement While Hearing Music: Effects of Structural Continuity on Feature Binding. *Music Perception*, *32*(1), 11–32. http://doi.org/10.1525/jams.2009.62.1.145.

Ehrlé, N., Samson, S., & Peretz, I. (2001). Normes pour un corpus musical. *Annee Psychologique*, *101*(4), 593–616. http://doi.org/10.3406/psy.2001.29569

Farooq, O., & Datta, S. (2001). Mel filter-like admissible wavelet packet structure for speech recognition. *IEEE Signal Processing Letters*, *8*(7), 196–198. http://doi.org/10.1109/97.928676

Fujioka, T., Trainor, L. J., & Ross, B. (2008). Simultaneous pitches are encoded separately in auditory cortex: an MMNm study. *Neuroreport*, *19*(3), 361–366. http://doi.org/10.1097/WNR.0b013e3282f51d91

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1-2), 103–138. http://doi.org/10.1016/0378-5955(90)90170-T

Gockel, H. E., Alsindi, S., Hardy, C., & Carlyon, R. P. (2017). Effect of Context on the Contribution of Individual Harmonics to Residue Pitch. *JARO - Journal of the Association for Research in Otolaryngology*, 1–11. http://doi.org/10.1007/s10162-017-0636-6

Graves, J. E., Micheyl, C., & Oxenham, A. J. (2014). Expectations for melodic contours transcend pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(6), 2338–47. http://doi.org/10.1037/a0038291

Houtsma, A., & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *The Journal of the Acoustical Society of America*, *87*(1), 304–310.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, *19*(1), 1–64.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Hutchins, S., Roquet, C., & Peretz, I. (2012). The vocal generosity effect: How bad can your singing be? *Music Perception*, *30*(2), 147–160. http://doi.org/10.1525/mp.2012.30.2.147

Hyde, K. L., & Peretz, I. (2004). Brains That Are out of Tune but in time. *Psychological Science*, *15*(5), 356–360.

IBM SPSS Statistics for Windows, Version 24.0. (2016). Armonk, NY: IBM Corp.

Kato, T., Omachi, S., & Aso, H. (2002). Asymmetric Gaussian and its application to pattern recognition. In *Proceedings of the Joint IAPR International Workshops SSPR 2002 and SPR 2002* (pp. 404–413).

Kessler, E., Hansen, C., & Shepard, R. (1984). Tonal schemata in the perception of music in Bali and in the West. *Music Perception*, *2*(2), 131–165.

Kirsch, I. (1985). Response expectancy as a determinant of experience and behavior. *American Psychologist*, 1189–1202.

Klapuri, A. (2008). Multipitch analysis of polyphonic music and speech signals using an auditory model. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(2), 255–266. http://doi.org/10.1109/TASL.2007.908129

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: “nonmusicians” are musical. *Journal of Cognitive Neuroscience*, *12*(3), 520–541. http://doi.org/10.1162/089892900562183

Krohn, K. I., Brattico, E., Välimäki, V., & Tervaniemi, M. (2007). Neural representations of the hierarchical scale pitch structure. *Music Perception*, *24*(3), 281–296.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology. Human Perception and Performance*, *18*(3), 739–751. http://doi.org/10.1037/0096-1523.18.3.739

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368. http://doi.org/10.1037/0033-295X.89.4.334

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 579–94.

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: cochlear nerve degeneration after “temporary” noise-induced hearing loss. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(45), 14077–85. http://doi.org/10.1523/JNEUROSCI.2845-09.2009

Larrouy-Maestri, P., Magis, D., Grabenhorst, M., & Morsomme, D. (2015). Layman versus professional musician: Who makes the better judge? *PLoS ONE*, *10*(8), 1–13. http://doi.org/10.1371/journal.pone.0135394

Larsen, E., Cedolin, L., & Delgutte, B. (2008). Pitch representations in the auditory nerve: two concurrent complex tones. *Journal of Neurophysiology*, *100*(3), 1301–1319. http://doi.org/10.1152/jn.01361.2007

Larson, S. (2004). Musical forces and melodic expectations: Comparing computer models and experimental results. *Music Perception*, *21*(4), 457–499.

Lau, B. K., Mehta, A. H., & Oxenham, A. J. (2017). Super-optimal perceptual integration suggests a place-based representation of pitch at high frequencies. *The Journal of Neuroscience*, *37*(37), 1507–17. http://doi.org/10.1523/JNEUROSCI.1507-17.2017

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423. http://doi.org/10.3758/BF03206733

Licklider, J. (1956). Auditory frequency analysis. In C. Cherry (Ed.), *Information Theory* (pp. 253–268). New York: Academic Press.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*, 5–55.

Lu, X., Sun, Y., Ho, H. T., & Thompson, W. F. (2017). Pitch contour impairment in congenital amusia: New insights from the Self-paced Audio-visual Contour Task (SACT). *PLoS ONE*, *12*(6), 1–15. http://doi.org/10.1371/journal.pone.0179252

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science*, *1*(4), 272–276. http://doi.org/10.1111/j.1467-9280.1990.tb00213.x

Marmel, F., Perrin, F., & Tillmann, B. (2011). Tonal expectations influence early pitch processing. *Journal of Cognitive Neuroscience*, *23*, 3095–3104. http://doi.org/10.1162/jocn.2011.21632

Marmel, F., Tillmann, B., & Dowling, W. (2008). Tonal expectations influence pitch perception. *Perception & Psychophysics*, *70*(5), 841–852. http://doi.org/10.3758/PP

McDermott, J. H., Keebler, M. V, Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: frequency resolution, not interval resolution, is special. *The Journal of the Acoustical Society of America*, *128*(4), 1943–51. http://doi.org/10.1121/1.3478785

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*, *19*(12), 1263–71. http://doi.org/10.1111/j.1467-9280.2008.02235.x

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, *20*(11), 1035–1041. http://doi.org/10.1016/j.cub.2010.04.019.Individual

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *The Journal of the Acoustical Society of America*, *102*(3), 1811–1820. http://doi.org/10.1121/1.420088

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, *48*(2), 169–178.

Micheyl, C., Bernstein, J. G. W., & Oxenham, A. J. (2006). Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise. *The Journal of the Acoustical Society of America*, *120*(3), 1493–1505. http://doi.org/10.1121/1.2221396

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1-2), 36–47. http://doi.org/10.1016/j.heares.2006.05.004

Micheyl, C., Keebler, M. V, & Oxenham, A. J. (2010). Pitch perception for mixtures of spectrally overlapping harmonic complex tones. *The Journal of the Acoustical Society of America*, *128*(1), 257–69. http://doi.org/10.1121/1.3372751

Monahan, C. B., Kendall, R. a, & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception & Psychophysics*, *41*(6), 576–600.

Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustic Society of America*, *54*(3), 610–619.

Moore, B. C. J. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *The Journal of the Acoustical Society of America*, *77*(5), 1853. http://doi.org/10.1121/1.391936

Moore, B. C. J., Glasberg, B. R., & Oxenham, A. J. (2012). Effects of pulsing of a target tone on the ability to hear it out in different types of complex sounds. *The Journal of the Acoustical Society of America*, *131*(4), 2927. http://doi.org/10.1121/1.3692243

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *The Journal of the Acoustical Society of America*, *80*(2), 479–83. http://doi.org/10.1121/1.394043

Moore, B. C. J., Glasberg, B. R., & Shailer, M. J. (1984). Frequency and Intensity Differences Limens for Harmonics within Complex Tones. *Journal of the Acoustical Society of America*, *75*(2), 550–561.

Moore, B. C. J., Huss, M., Vickers, D. a, Glasberg, B. R., & Alcántara, J. I. (2000). A test for the diagnosis of dead regions in the cochlea. *British Journal of Audiology*, *34*(4), 205–224. http://doi.org/10.3109/03005364000000131

Moreau, P., Jolicœur, P., & Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials. *Brain and Cognition*, *81*(3), 337–344. http://doi.org/10.1016/j.bandc.2013.01.004

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

Noll, A. M. (1967). Cepstrum pitch determination. *The Journal of the Acoustical Society of America*, *41*(2), 293–309. http://doi.org/10.1121/1.1910339

Ohm, G. S. (1843). Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen (On the definition of a tone and related theory of a siren and similar tone-producing devices). *Annalen Der Physik Und Chemie*, *59*, 513–565.

Oxenham, A. J., Bernstein, J. G. W., & Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(5), 1421–1425. http://doi.org/10.1073/pnas.0306958101

Oxenham, A. J., Micheyl, C., Keebler, M. V, Loper, A., & Santurette, S. (2011). Pitch perception beyond the traditional existence region of pitch. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(18), 7629–34. http://doi.org/10.1073/pnas.1015291108

Palmer, C., & Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Perception & Psychophysics*, *56*(3), 301–312. http://doi.org/10.3758/BF03209764

Parncutt, R., & Bregman, A. S. (2000). Tone profiles following short chord progessions: Top-down or bottom-up? *Music Perception*, *18*(1), 25–57.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, *72*(6), 1788–1803. http://doi.org/10.1121/1.388652

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: the influence of context and learning. *Music Perception*, *23*(5), 377–405.

Pelofi, C., de Gardelle, V., Egré, P., & Pressnitzer, D. (2017). Interindividual variability in auditory scene analysis revealed by confidence judgements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160107. http://doi.org/10.1098/rstb.2016.0107

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain : A Journal of Neurology*, *132*(Pt 5), 1277–86.

http://doi.org/10.1093/brain/awp055

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varietes of Musical Disorders: The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(5), 976–986. http://doi.org/10.1037/0096-1523.20.5.976

Plack, C. J., Oxenham, A. J., Fay, R. R., & Popper, A. N. (Eds.). (2005). *Pitch: Neural Coding and Perception*. *Spinger Handbook of Auditory Research* (Vol. 24). New York: Springer Science & Business Media. http://doi.org/10.1007/0-387-28958-5

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*(1), 1–11. http://doi.org/10.1016/j.cognition.2004.09.008

Plomp, R. (1967). Pitch of Complex Tones. *The Journal of the Acoustical Society of America*, *41*(6), 1526. http://doi.org/10.1121/1.1910515

Plomp, R., & Levelt, W. J. (1965). Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America*, *38*(4), 548–60.

Pressnitzer, D., Patterson, R. D., & Krumbholz, K. (2001). The lower limit of melodic pitch. *The Journal of the Acoustical Society of America*, *109*(5), 2074–2084. http://doi.org/10.1121/1.1359797

Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. *Psychology of Music*, *18*(1), 60–72. http://doi.org/10.1177/0305735690181005

Reiss, L. a J., Turner, C. W., Karsten, S. a, & Gantz, B. J. (2013). Plasticity in Human Pitch Perception Induced by Tonotopically Mismatched Electro-Acoustic Stimulation. *Neuroscience*, *256*, 43–52. http://doi.org/10.1016/j.neuroscience.2013.10.024

Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. *The Journal of the Acoustical Society of America*, *42*(1), 191–198. http://doi.org/10.1121/1.1942972

Russo, F. a, Ives, D. T., Goy, H., Pichora-Fuller, M. K., & Patterson, R. D. (2012). Age-related difference in melodic pitch perception is probably mediated by temporal processing: empirical and computational evidence. *Ear and Hearing*, *33*(2), 177–86. http://doi.org/10.1097/AUD.0b013e318233acee

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. http://doi.org/10.1016/S0010-0277(98)00075-4

Santurette, S., & Dau, T. (2011). The role of temporal fine structure information for the low pitch of high-frequency complex tones. *The Journal of the Acoustical Society of America*, *129*(1), 282–292. http://doi.org/10.1121/1.3518718

Santurette, S., Dau, T., & Oxenham, A. J. (2012). On the possibility of a place code for the low pitch of high-frequency complex tones. *The Journal of the Acoustical Society of America*, *132*(6), 3883–95. http://doi.org/10.1121/1.4764897

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition*, *58*(1), 75–125.

Schellenberg, E. G. (1997). Simplifying the Implication-Realization model of Melodic

Expectancy. *Music Perception*, *14*(3), 295–318.

Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: Tests of children and adults. *Journal of Experimental Psychology: General*, *131*(4), 511–537. http://doi.org/10.1037//0096-3445.131.4.511

Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Memory & Cognition*, *43*(7), 1021–1031. http://doi.org/10.3758/s13421-015-0519-1

Schenker, H. (1935). *New musical theories and fantasies*. (J. Rothgeb, Ed.). New York: Schirmer Books.

Schmuckler, M. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception*, *7*, 109–150.

Schoenberg, A. (1911). *Harmonielehre*. Leipzig and Vienna: Verlagseigentum der Universal-Edition.

Schroeder, M. R. (1970). Synthesis of Low-Peak-Factor Signals and Binary Sequences with Low Autocorrelation. *IEEE Transactions on Information Theory*, *16*(1), 85–89. http://doi.org/10.1109/TIT.1970.1054411

Seebeck, A. (1841). Beobachtungen über einige Bedingungen der Entstehung von Tönen (Observations on some conditions of tone formation). *Annalen Der Physik Und Chemie*, *53*, 417–436. http://doi.org/10.1002/andp.18411290702

Shackleton, T. M., & Carlyon, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *The Journal of the Acoustical Society of America*, *95*(6), 3529–3540. http://doi.org/10.1121/1.409970

Shera, C. A., Guinan, J. J., & Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements.

Shinn-Cunningham, B. G., Lee, A. K. C., & Oxenham, A. J. (2007). A sound element gets lost in perceptual competition. *Proceedings of the National Academy of Sciences*, *104*(29), 12223–12227. http://doi.org/10.1073/pnas.0704641104

Siegel, R. J. (1965). A replication of the mel scale of pitch. *The American Journal of Psychology*, *78*(4), 615. http://doi.org/10.2307/1420924

Song, X., Osmanski, M. S., Guo, Y., & Wang, X. (2015). Complex pitch perception mechanisms are shared by humans and a New World monkey. *Proceedings of the National Academy of Sciences*, *2015*(15), 201516120. http://doi.org/10.1073/pnas.1516120113

Stevens, J., & Hall, J. (1966). Brightness and loudness as functions of stimulus duration. *Perception & Psychophysics*, *1*, 319–327.

Stevens, S. S. (1935). The relation of pitch to intensity. *The Journal of the Acoustical Society of America*, *6*, 150–154. http://doi.org/10.1121/1.1902092

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, *8*(3), 185–190.

Studebaker G.A. (1985). A "rationalized" arcsine transform. *Journal of Speech and Hearing Research*, *28*, 455–462. http://doi.org/10.1044/jshr.2803.455

Szpunar, K. K., Schellenberg, E. G., & Pliner, P. (2004). Liking and memory for musical stimuli as a function of exposure. *Journal of Experimental Psychology: Learning,* *Memory, and Cognition*, *30*(2), 370–81. http://doi.org/10.1037/0278-7393.30.2.370

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science*, *32*(2), 418–44. http://doi.org/10.1080/03640210701864089

Thompson, W. F., Balkwill, L. L., & Vernescu, R. (2000). Expectancies generated by recent exposure to melodic sequences. *Memory & Cognition*, *28*(4), 547–55.

Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia: Evidence from a gating paradigm. *Cortex*, *59*, 84–94. http://doi.org/10.1016/j.cortex.2014.07.012

Tillmann, B., Dowling, W. J., Lalitte, P., Molin, P., Schulze, K., Poulin-Charronnat, B., … Bigand, E. (2013). Influence of Expressive Versus Mechanical Musical Performance on Short-term Memory for Musical Excerpts. *Music Perception: An Interdisciplinary Journal*, *30*(4), 419–425. http://doi.org/10.1525/mp.2013.30.4.419

Tillmann, B., Janata, P., Birk, J., & Bharucha, J. J. (2008). Tonal centers and expectancy: facilitation or inhibition of chords at the top of the harmonic hierarchy? *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1031–1043. http://doi.org/10.1037/0096-1523.34.4.1031

Tillmann, B., & Marmel, F. (2013). Musical expectations within chord sequences: Facilitation due to tonal stability without closure effects. *Psychomusicology: Music, Mind, and Brain*, *23*(1), 1–5. http://doi.org/10.1037/a0030454

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition*, *71*(3), 259–264. http://doi.org/10.1016/j.bandc.2009.08.003

Trainor, L. J., Marie, C., Bruce, I. C., & Bidelman, G. M. (2014). Explaining the high voice superiority effect in polyphonic music: Evidence from cortical evoked potentials and peripheral auditory models. *Hearing Research*, *308*, 60–70. http://doi.org/10.1016/j.heares.2013.07.014

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants’ and adults' sensitivity to western musical structure. *Journal of Experimental Psychology. Human Perception and Performance*, *18*(2), 394–402. http://doi.org/10.1037/0096-1523.18.2.394

Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants’ and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 965–75.

Verschuure, J., & van Meeteren, A. A. (1975). The effect of intensity on pitch. *Acustica*, *32*, 33–44.

von Hippel, P., & Huron, D. (2000). Why Do Skips Precede Reversals ? The Effect of Tessitura on Melodic Structure. *Music Perception*, *18*(1), 59–85.

Vuvan, D. T., & Schmuckler, M. a. (2011). Tonal hierarchy representations in auditory imagery. *Memory & Cognition*, *39*(3), 477–490. http://doi.org/10.3758/s13421-010-0032-5

Wang, J., Baer, T., Glasberg, B., Stone, M., Datian, Y., & Moore, B. C. J. (2012). Pitch perception of concurrent harmonic tones with overlapping spectra. *The Journal of the Acoustical Society of America*, *132*(1), 339–56. http://doi.org/10.1121/1.4728165

Wapnick, J., Bourassa, G., & Sampson, J. (1982). The perception of tonal intervals in isolation and in melodic context. *Psychomusicology*, *2*(1), 21–37. http://doi.org/10.1037/h0094264

Warren, R., Obusek, C., Farmer, R., & Warren, R. (1968). Auditory Sequence : Confusion of Patterns Other Than Speech or Music. *Science*, *164*(3879), 586–587.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, *64*(2), 198–207.

Whiteford, K. L., & Oxenham, A. J. (2015). Using individual differences to test the role of temporal and place cues in coding frequency modulation. *The Journal of the Acoustical Society of America*, *138*(5), 3093–3104. http://doi.org/10.1121/1.4935018

Whiteford, K. L., & Oxenham, A. J. (2017). Auditory deficits in amusia extend beyond poor pitch perception. *Neuropsychologia*, *99*(October 2016), 213–224. http://doi.org/10.1016/j.neuropsychologia.2017.03.018

Yeh, C., Roebel, A., & Rodet, X. (2010). Multiple fundamental frequency estimation and polyphony inference of polyphonic music signals. *IEEE Transactions on Audio, Speech and Language Processing*, *18*(6), 1116–1126. http://doi.org/10.1109/TASL.2009.2030006

Yin, P., Fritz, J. B., & Shamma, S. A. (2010). Do ferrets perceive relative pitch? *The Journal of the Acoustical Society of America*, *127*(3), 1673–1680. http://doi.org/10.1121/1.3290988

Zarate, J., Ritson, C., & Poeppel, D. (2012). Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary? *The Journal of the Acoustical Society of America*, *132*(2), 984–93. http://doi.org/10.1121/1.4733535

Zilany, M., Bruce, I., & Carney, L. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical …*, *135*(1), 283–6. http://doi.org/10.1121/1.4837815