Ventriloquial Dummy Tones: Embodied Cognition of Pitch Direction

JOHN GRANZOW
Bachelor of Arts, University of Lethbridge, 1999

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Psychology
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

To my Father

> Hearing is by means of the ears, because within them is an empty space, and this empty space resounds

Alcmaeon of Crotona

> To it's credit, Rutherford's two page theory was parsimonious, to its discredit it just shoved the problem one stage up

Alain De Cheveigne

> In a world which is characterised not by the autonomization of sensory channels and their corresponding media but by their remorseless interchange, the voice is kind of a sluggish impediment to the logic of sensory conversion and commutation

Steven Connor

> As all action is by its nature to be figured as extended in breadth and in depth, as well as in length; and spreads abroad on all hands...as well as advances towards completion - so all narrative is, by its nature, of only one dimension, only travels forward towards one, or towards successive points; narrative is linear, action is solid. Alas for our chains and chainlets, of "causes and effects", which we so assiduously track through certain handbreadths of years and square miles, when the whole is a broad deep immensity.

Thomas Carlyle, On History

> ...most sound sources are quite small, and therefore the sound is produced in the form of a spherical wave, in which sound travels out from the source in every direction.

Dr. Tempest

> ”The sea was not a mask. No more was she.
> The song and water were not medleyed sound
> Even if what she sang was what she heard.
> Since what she sang was uttered word by word.
> It may be that in all her phrases stirred
> The grinding water and the gasping wind;
> But it was she and not the sea we heard.

Wallace Stevens

> When articulation and sound wave go their separate ways, which way does perception go?

Alvin Liberman

Abstract

Tone pairs constructed with the frequencies of the overtones moving in opposition to the missing fundamental frequencies they imply, produce expertise differences in the tracking of pitch direction. One interpretation of this result is that it arises as a function of rudimentary differences in the perceptual systems of musicians and non-musicians. Several experiments suggest instead a more embodied source of expertise to be found in vocal mediation such that the effect of musical experience in these tasks is the result of the most salient action of musicians: making sound.

Acknowledgments

I would to thank my supervisor Professor John R. Vokey, who never lets the present zeitgeist hamper scientific discovery and offers lucky students such as me a relentless passion for knowledge, and collaboration. I was also very fortunate to have such a knowledgeable and engaged committee; Professor Louise Barrett kept my shelf heavy with the ballast of good books; Professor Drew Rendall offered his expertise whenever our study wandered into the linguistic domain; and Professor Matt Tata's experience in auditory research was much help when working out the structure of these experiments. Thanks as well to Professor Scott Allen, who sat in on some of my committee meetings and gave useful feedback. Thanks to Leanne Wehlage for helping me get participants. Thanks as well to my lab mates, Shelley Gross, Michelle Corcoran, Mitch LaPointe, Tom Rutherford and the wider cohort of graduate students that have made life at the university fun.

I have dedicated this thesis to my father Carl, whose encouragement, support, laughter and guidance will always be remembered and cherished. I thank my Mother Friederike for her strength, love and support through this time. I thank my brother Michael; although his bass playing delayed the completion of the thesis his emergency delivery of bell jars made it all possible. Thank you to my wonderful sisters Kara and Andrea for their constant and continuing support, and to their children Henry, Suzana, Annika, and Samuel who pulled me from the intertia of life-in-a-chair with their surprise visits. Thank you to Madeleine Baldwin, who put up with two grumpy grad students through the grind. I thank my good friend Christine Johnson for her constant support and contagious laughter, an antidote to ever taking oneself too seriously. And finally thanks to my life time friend Denton Fredrickson for the beginnings of experimental life in grade eleven spare, and for keeping me active ever since in the scrap booking community.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Perceptual Dimensions

The science of psychoacoustics inherits from music theory categories used to describe organized sound: Timbre, rhythm and meter, harmony, pitch. Empirical studies reveal the nuances of these perceptual dimensions as well as how they might interact.

The category of pitch has received the most attention due in part to its importance in musical scale systems and vowel perception in language. Patel (2008) proposes that pitch exceeds other auditory dimensions in the possibilities it affords for organizing sound due to its duplex cognitive representation in both height and chroma (Shepard, 1982) (see Figure 1.1). Height is the vertical metaphor along which it is said we map the rate of vibration per second of acoustic waves. Pitch perception rises as this overall rate of vibration increases, and the relationship is on a logarithmic scale. Chroma refers to the percept of relationships between notes, the most notable being the unison or octave where an equivalence is perceived when the frequency approaches one of its multiples (see Ueda & Ohgushi, 1987, for a review of this three-dimensional representation of pitch).

With these manifold affordances, pitch shifts, rather than some other dimension such as timbre or loudness, is the dimension along which sound is predominantly organized in music.

1.2 Pitch

The definition of musical pitch taken from the American National Standards Institute begins in a round-about way: "That attribute of auditory sensation in terms of which sounds

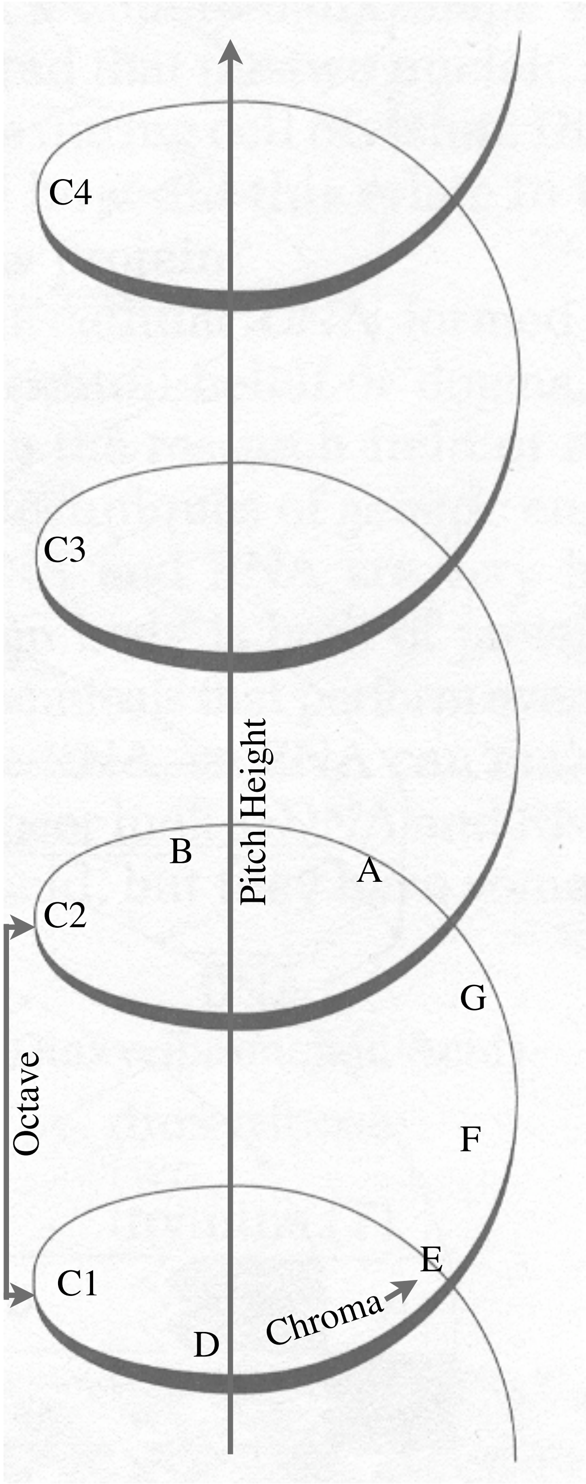


Figure 1.1: Height and chroma: Two dimensions of pitch perception

may be ordered on a scale extending from low to high"(ANSI, 1994, p. 34). This definition coheres to the circularity Locke (1690/1975) predicted in descriptions of what he called *secondary properties* or properties that are induced only in the mental. Accordingly, the ANSI definition begins by simply invoking the names we assign pitch in relative metaphorical space. This mapping is most often used as a contrast-dependent description (i.e., tone y is higher than tone x) and as such illustrates the relative nature of pitch perception for the majority of people. Aside from the few who retain categorical pitch perception (authentic pitch), most people evaluate the location of pitch on a scale by establishing its contrast to neighbouring notes. These differences are conserved between different octaves. As long as the relative pitch differences are retained, a melody can be recognized regardless of the initial note (Attneave & Olson, 1971).

This spatial metaphor of height is culturally specific; to grow up in in Bali or Java is to describe a pitch as shrinking, rather than rising, and in Ancient Greece, a falling pitch was thought, rather, to gain weight (Zbikowski, 2002). In what might seem an incommensurable difference in metaphorical mappings across cultures, the uses of these cross-domain mappings are indicative of a common ground: Pitch, like other subjective percepts, cannot easily be defined as an aggregate of constitutive events or sequences and so it is described with recourse to parallel visual metaphors.

Furthermore, in the operation of such metaphors, Lakoff (1989) has proposed that something might be preserved about perception across cultural variants. For example, the fact that, in surveys, the Balinese guess correctly the direction in which the height metaphor operates when applied to pitch, and those surveyed from western musical backgrounds guess correctly whether a high note would, in Bali, be considered small or large, seems to imply some common operation of these metaphors (M. Turner, 1990; Zbikowski, 2002). In short, where the diversity of descriptive conventions suggest relative differences, the between framework invariants invite inquiry into universal cognitive patterns of perception that these metaphors cue.

The second part of this standard definition includes a succinct list of correlates between the physical properties of acoustic waveforms and the perceptual experience: “Pitch depends primarily on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus” (ANSI, 1994, p. 34). The two parts of the definition now exemplify the distinction that Locke made between aforementioned *secondary properties* and *primary properties* which are properties inherent to the object itself. In general, pitch perception research has been preoccupied with moving from the circularity of the first part of the definition, to the linear framework of the second part, in order to provide an account of the causal chain between acoustic properties, anatomical constraints and higher order auditory percepts. This research conforms to the *causal methodology* as outlined by Levitin (2002), where confounding variables are controlled in order to observe and draw inferences from the relationship between independent variable (in this case the waveform of the stimulus) and dependent variable (response pattern of the participant). For this type of research, the discovery that has most influenced the way we manipulate this independent variable was Fourier’s (1820) transform.

1.3 Analyses of Sound

The most influential insight in the field of acoustics as it pertains to pitch perception was the discovery that sounds emanating from one source can have a multitude of constituent frequencies. Cheveigné (2004), in his review of pitch perception models, traces this insight back to ancient Greece. However, it was not until Fourier (1820) devised his famous transform that the mathematical proof was in place. Fourier (1820) demonstrated that any waveform could be mathematically deconstructed into a series of constituent sinusoidal pure frequencies. From such methods applied to acoustic analyses, reliable predictions can be made as to what pattern of frequencies will give rise to the perception of pitch. Fourier (1820) thus provided an invaluable tool to explore the underlying assumption of bottom-up models of auditory perception: the information necessary to understand perception is assumed to exist in the physical description of the stimulus. The pattern in the acoustic waveform that most reliably induces pitch is called the *Harmonic Series*.

1.4 Harmonic Series

Objects such as pipes, strings and vocal chords tend to produce regular patterns of vibration in the air. When several of these vibrations oscillate together such that those at higher frequencies occur at integer multiples of the lowest one, the series is called harmonic and this temporal regularity is called periodicity.

In accordance with bottom-up pitch perception models, to hear a complex tone as pitched, the auditory system must resolve these multiple vibrations into a unitary sensation. The singularity of a harmonic sound source, (a string, a voice), potentially lost in the multitude of frequencies to which it gives rise, is reconstituted in perception allowing the listener to interpret and enumerate what is making sound in the environment.

The harmonic series can also be reinforced with several sources (multiple stringed instruments or a chorus of voices sounding with between-tone frequencies matching within-tone harmonics). These related sounds can give rise to the impression that many sound sources are one. Such events exploit the perceptual tendency to fuse concurrent vibrations that are harmonically related. In both speech and music, such fusion depends on many other factors such as attack and onset synchrony between sounds (Denbigh & Zhao, 1992; Huron, 2001).

Within this harmonic pattern of vibration, Sauveur (1701) was the first to make the distinction within the spectrum between the “fundamental” (F_0) and the “overtones”. The

F_0 is the lowest frequency to which the overtones are related as integer multiples. Both the period of cycling (time domain) and the frequency components of the fundamental and overtones (spectral domain) have been explored as causal antecedents to the sensation of pitch. Energy at the F_0 was the prime candidate for the cause of pitch as its rise induces a concomitant rise in the percept. Thus, historically, pitch has been associated with the presence of energy at the F_0 .

Fission also occurs; in these cases overtones in a harmonic pattern can be heard out as separate components. Sano and Jenkins (1991) referred to this as analytic pitch perception to distinguish it from the synthetic percept of pitch. Bregman (1994) preferred the terms "partial pitch" and distinguished this percept from "global pitch" where a single tone is heard.

Tuvan and Inuit throat singing employ vocal techniques to emphasize specific overtones such that they can be segregated as separate tones by the listener. Such musical traditions depend on our perceptual ability to attend to overtones, reversing in some sense the process of fusion that might underly pitch perception. In general, attentive listening can reveal the presence of discrete overtones in many harmonic sounds. Mersenne (1636) claimed he could hear up to five overtones as pitched in a vibrating string. This possibility to focus our hearing in order to isolate perceptually component frequencies of an otherwise singular percept, already suggests the need to consider listening as an intentional action that varies the percept from the top down.

Sounds that deviate from this wave pattern are called inharmonic and consequently are not classified as having pitch. The basis for the classification of tones as "non-pitched" is that they cannot be placed with any precision on a musical scale. Yet pitch has been shown to have varying strengths and, when we compare two sounds otherwise considered non-pitched (e.g., bass, drum and triangle), we can often distinguish the sounds on the basis of height, suggesting that, like other perceptual domains, pitch exists on a continuum

(Rakowski, 2004).

1.5 Timbre

The harmonic overtone series is perfectly correlated (except for amplitude) with the frequency of the fundamental and therefore pitch. Nonetheless, this series, as distinct from the fundamental, has been considered to influence another perceptual dimension: timbre. Timbre is the quality of the sound through which we can distinguish, for example, a brass instrument from a woodwind, even when they are producing the same pitch or note at the same duration and loudness. The series of the overtones and their relative amplitudes strongly influence these timbral effects and vary among musical instruments. In defining timbre through the process of holding pitch constant, their potential independence is foregrounded. Furthermore, we might assign timbre and pitch independent causal sources in the components of the waveform (i.e., pitch from the F_0 and timbre from the amplitude pattern of overtones).

Principally through the work of McAdams, Winsberg, Donnadieu, De Soete, and Krimphoff (2004), timbre has been shown to be a multidimensional percept. Furthermore, the literature now suggests that timbre and pitch are less independent than previous orthogonal definitions imply (see Krumhansl & Iverson, 1992, for a review of the interaction between musical pitch and timbre).

1.6 Causes in Acoustic Properties

The attempt to observe elegant correspondences between pitch perception and the physical properties of the stimulus is a very old enterprise. Having none of the present tools of acoustic analyses, early experiments of pitch explored relationships between, rather than within, sound stimuli. Pythagorus attempted to map perception monotonically to the simple integer ratios that he observed on his monochord, proposing that our perceptions reflected this mathematical simplicity (i.e., 1:2 (octave), 2:3 (fifth), 3:4 (fourth)). Aristoxenus, in contrast, focused on the qualitative aspects of what is heard (Macran, 1902). For him, the idiosyncrasies of auditory perception were to be the measure of the musical scale, not the invocation of mathematical elegance.

This debate recurs: In the early 17th century Mersenne (1636) also looked for perceptual correspondences in mathematics . This time it was his colleague Rènè Descartes who provided the critique of such a notion (Cheveigné, 2004). Presently, there is considerable evidence that there are no simple isomorphisms between perception and a sound's mathematical description. If there were, we might expect differing discrimination depending on the presence or absence of intervals that correspond to simple integer ratios. Categorical perception might reflect such a correspondence if indeed percepts reflected simple ratios. Burns and Ward (1978) have demonstrated, however, that, with the exception of the octave, interval discrimination does not show perceptual biases toward simple ratios.

Analysis techniques emerging from Fourier's (1820) theorem allowed researchers to ask a more focused question: What within the waveform itself might act as the physical cause of pitch? Through spectral analyses, we can now observe the components that correspond to Sauveur's (1701) distinction between 'fundamental' and 'overtones', and these component frequencies can be manipulated as variables in order to infer causes and interactions between waveform and percept. However, when Seebeck (1841) demonstrated that energy at the frequency of the F_0 was not necessary to hear pitch, the project of mapping pitch perception monotonically to acoustic waveform was brought into question.

1.7 The Missing Fundamental

The F_0 is the frequency that corresponds to the overall wavelength of a harmonic sound. As such, it often is the frequency at which there is the most energy, as it is reinforced by the harmonically related partials. The F_0 was historically thought to be the physical correlate, and, by implication, most likely cause of pitch perception; as it increased or decreased in frequency, the perception of pitch would rise or fall, respectively. This notion that pitch is associated with the F_0 culminated in Ohm's (1843) acoustical law that was subsequently supported by Helmholtz (1836/1954) and stated that energy at the frequency of the fundamental was required for the perception of pitch.

The main contention of Ohm's (1843) acoustical law was already in question in the demonstrations of Seebeck (1841) who created pure-tone sirens that in combination were used to reproduce overtone series in the absence of the fundamental frequency. He noticed that the pitch associated with this now missing fundamental was still heard. Yet Helmholtz (1836/1954) reiterated Ohm's (1843) assumption of requisite energy at the frequency of the F_0, and coined the term "Ohm's law" that has had an enduring influence on hearing models (Plack & Oxenham, 2005; R. Turner, 1977). In the twentieth century, the issue was taken up again when F. Schouten (1938) used synthesized tones to confirm that the perception of pitch associated with a given F_0 did not depend on the presence of energy at that frequency; for pitch perception of a low (and missing) frequency, all that was required was a subset of the associated overtone series.

The insight that pitch perception can be induced without the F_0 led researchers to manipulate the overtone series in order to observe the limits and idiosyncrasies of the auditory system as it resolves multi-frequency sounds into a singular pitch sensation. For example, in the literature exploring musical expertise, varying the overtone series interfered with the capacity of non-musicians to discriminate pitch differences, whereas musicians were less likely to be influenced by these manipulations (see Krumhansl & Iverson, 1992; Pitt, 1994; Warrier & Zatorre, 2002).

The missing fundamental illusion goes by many other names: virtual pitch, residual pitch, periodicity pitch and low pitch. The proliferation of names is the result of a history of rivalling models that aim to account for this auditory phenomenon. Hereafter the illusion (a misnomer because pitch is already in some sense an illusion, and it is uncertain what an illusion within an illusion might mean) will be called the implied F_0 .

1.8 Causes in the Ear

Unlike the percept for which they try to account, models for an objective description of pitch perception are still far from unified (Lee, 2004). Pitch constancy through tones both with and without the F_0 at least reinforces the notion that pitch is a construction or inference of the auditory system in the presence of general patterns. Therefore, the mechanics of that system are explored for the source of the organic "inference" of pitch.

1.8.1 Place Theory and Pattern Matching

Emerging from Du Verney's (1683) theory of resonance in the ear, the place theory holds that pitch perception arises from a sensitivity to the location of maximum vibration on the basilar membrane. When sound energy is transmitted to the cochlear fluid, vibrations excite the hair cells at the location along the basilar membrane that correspond to the period frequency of the stimulus or F_0 .

Pitch is encoded through the selective response to physical vibrations in the air. This representation of pitch associated with "place" is observed in both the middle ear (Ruggero, 1992) and further up the pathway in the auditory cortex (Kaas & Hackett, 1999). The use of pure tones (tones with one sinusoidal frequency) in hearing research made such an account intuitive, as one frequency could activate a single cluster of hair cells on basilar membrane. The implied F_0 revealed the potential inadequacy of this model as a unitary percept of pitch occurred in the absence of energy at the corresponding "place" of the F_0 .

Fourier's (1820) transform had another profound effect on pitch perception models as it was assumed that such mathematical elegance must be a functional analogue to the operation of the ear in the presence of harmonic sound (Cheveigné, 2004). Accordingly, early models, such as the place theory, may have been influenced by this suspicion that what was going on in the ear must be something roughly like an organic Fourier analysis. This perspective on the ear as a kind of electrochemical frequency analyzer became the basis for the place model of pitch perception firmly established in the seminal work of Helmholtz (1836/1954).

To account for the phenomenon of implied F_0 , Helmholtz (1836/1954) established a theory where the ear was akin to an assymetrical vibrator that generated energy at the frequency corresponding to the difference between two partials. This idea led Helmholtz (1836/1954) to posit that such endogenous emissions might restore energy at the F_0 when it was not present in the stimulus, a notion that would explain implied pitch.

More recently, the non-linear response of the cochlea (Yates, Winter, & Robertson, 1990) has been shown indeed to produce distortions that correspond to the frequency differences between the presented sinusoids (Pressnitzer, Patterson, & Krumholz, 2001). In a harmonic tone, this difference will be the F_0 when the partials are contiguous, and it is, therefore, possible for distortions to activate the cochlea at the place associated with the frequency of the F_0 (Plack & Oxenham, 2005). Although evidence accumulates for cochlear distortion effects, Licklider's (1956) demonstration using a masking techniques serves to confirm that a wholly different kind of pitch perception is at work with the implied F_0 . Licklider (1956) used a masking technique to demonstrate that F_0 percepts endured in the presence of a masking noise that cancels any otoacoustic emissions at the frequency of the implied F_0 .

Place theories were the precursors of pattern matching accounts of pitch perception (Goldstein, 1973; Wightman, 1973). One of the most notable of these is Terhardt's (1974) place theory where learning accounts for the perception of virtual pitch. Virtual pitch was Terhardt's term for the extraction of pitch from complex tones as distinct from the spectral pitch of pure tones. In this model, exposure to harmonic patterns induces something analogous to a statistical algorithm for extracting pitch. Listeners learn to correlate spectral cues in order to infer the implied pitch. One thing the invocation of learning implies is that we might find expertise effects among those who engaged in ear training, a topic that will be explored in following chapters. This modelling of the ear through statistical algorithms can become exceedingly removed from physiological evidence of auditory mechanics. Such models are useful, nonetheless, as analogies for a perceptual system that seems to do something of the kind. For example, Cheveigné's (2004) recent pattern matching algorithm presents a detailed algorithm for extracting the F_0, but also exemplifies to what remove such models are from any account of how such an algorithm would operate in an organic system. In brief, Cheveigné's (2004) algorithm is the accumulation of information in histogram bins that represent sub-harmonics of each partial in a complex sound. Because all harmonics will be divisible by the F_0, the histogram associated with the F_0, as well as an infinite number of sub-harmonics below the F_0 will be the highest. The system would only have to choose the highest frequency with the highest histogram to determine the F_0.

1.8.2 Temporal Models and Autocorrelation

A longstanding alternative to place theory of pitch perception has been a system sensitive to the time domain of periodic waveforms. Such a system would depend on neural activity synchronous with the time difference between the peaks of a periodic wave as it cycles. Such models are limited to frequencies whose periods do not exceed the refractory period of implicated neurons in the auditory system as well as phase locking limitations of neurons (Stainsby & Cross, 2009). The temporal theory assumed some kind of counting mechanism whereby the rate of the periodic tone could be calculated. Furthermore there was no consensus on exactly what was being counted, with Boer (1976) proposing peaks and J. Schouten, Ritsma, and Cardozo (1962) proposing other "landmarks" in the cycling pattern. As both the place and temporal models of pitch perception have explanatory power and inherent limits, a combination has been proposed where high pitches are encoded by place and lower frequencies by time (Gockel, Moore, & Carlyon, 2001; Moore, 2003). Common to these models is a rudimentary, physical and physiological level of investigation.

1.9 The Brain

Stewart (2008) outlines a history of seeking anatomical correlates to perceptual/behavioral variance (i.e., musicians vs nonmusicians), proceeding from simple (though potentially apocryphal) observations of Mozart's unusual pinna, to Auerbach's early equation of temporal and parietal differences with musical ability (Auerbach, 1906), and finally to contemporary neuroimaging of the cortex of musicians.

Using present techniques to image neural actitivity in the brain, Zarate (2008) associated musicians' capacity to resist shifted feedback in singing tasks with differing neural activity; Schlaug, Andrea, Overy, and Winner (2005) observed structural changes in the brain among children after only fifteen months of musical training. Among musicians, (Schneider et al., 2002) observed 130 percent increase in the size of the Heschl gyrus compared to those without musical training. These are only a few of numerous studies revealing anatomical correlates with musical training and ability.

Of course, imaging techniques have come a long way since the mere inspection of Mozart's missing tragus, yet the orders of magnitude between perception and neural activity still pose methodological problems to inferences of perceptual differences from physiological ones.

Compared to behavioral data, structural and functional differences in musicians' brains (Gaser & Schlaug, 2003) comprise a vastly different level of observation. Nonetheless, studies revealing expertise effects in conscious responses to acoustic stimuli lean on these anatomical data (when they corroborate) as part of the accumulating evidence for expertise.

It seems with compounding evidence for musical brains and musical responses we might safely surround the black box of perception and say that, as a consequence, it too, must be different. Yet, in the laboratory, attempts to infer the quality of the sensorium from response patterns, or brain images can be tricky. In the present series of studies, it becomes clear that musicians in the context of a recent pitch perception test might differ in performance from nonmusicians not as a function of acute hearing differences but more general musical/ behavioural ones that present paradigms overlook.

1.10 Expertise

Musical expertise is associated with a heightened sensitivity to pitch. In psychophysical tests, for example, musicians show a smaller frequency difference limen across synthetic and natural tones indicating increased pitch sensitivity (Meyer, 1978). Furthermore, musical training predicts the ability to resolve and track pitch contour more reliably when the F_0 is missing and other disorienting patterns are introduced into the spectrum (Seither-Preisler et al., 2007).

In general, listeners with no musical training can discriminate intervals of about 80 cents; this interval happens to fall below the smallest relevant interval of western music: one semitone (in equal temperament tuning, the interval of one semitone equals 100 cents) (Burns E & Ward, 1978). In interval detection tasks, participants are asked to compare intervals that start at different frequencies and subsequently report which one is larger. This perceptual ability is called interval perception and is distinct in the literature, [and, according to that literature, distinct in neural provenance (Liegeois-Chauvel, Peretz, Babai, Laguiton, & Chauvel, 1998; Schuppert, Munte, Wieringa, & Altenmüller, 2000)], from contour perception which is the much less complex task of tracking relative heights.

Contour perception can be represented with a binary system where, for example, the notes are simply notated by their height relative to the previous note (+ + - would notate the second, third, and fourth notes of ba ba black sheep). Stalinski, Schellenberg, and Trehub (2008) have recently confirmed that participants show a much higher sensitivity to contour, with adults detecting shifts of 0.1 semitones.

Given such highly precise percepts, any expertise effects demonstrated at contour differences greater than one semitone would be surprising. Such small intervals are even avoided in musical systems that could conceivably contain them (Patel, 2008). The only documented cross-cultural exception where intervals under a semitone are played is in Malenesia where a 33 cent interval was observed (Zemp, 1981). Not surprisingly, the intervals defining musical systems seem to exceed the bandwidth of cochlear and higher order discrimination.

Demonstrations of expertise at fine-grained thresholds are often more technical or perceptual than musical (Sloboda, 1992). Sloboda (1992) stresses that these psychometric expertise effects are not to be confused with the more species-wide musical expertise associated with an emotion-structure link between acoustical events and perception. It is this link that makes music a universal human activity. Furthermore, it is this general capacity that makes the use of music appealing for studies that hope to demonstrate general cognitive processes. A tension emerges between this general human expertise and the fine grained sensory ones produced by psychometric testing through which distinctions between musicians and nonmusicians might be ratified.

If these sensory musical expertise effects are equated narrowly with musical expertise, it can render music, like other aptitudes, synonymous with a battery of refined tests for its detection (aptitude categories and their tests are prone to this classic chicken/egg problem). Although the results of these tests may be highly correlated with musical ability, they may also reify musical expertise on the basis of perceptual thresholds that are not wholly relevant for the activity of interest. It has been shown that even such rare abilities as authentic pitch, an ability that would seem to have obvious advantages to musical activity, cannot be shown to predict musicality (Patel, 2008). Sensory expertise effects, then, are most strikingly demonstrated in tasks that require a degree of sensitivity exceeding what might be observed, or necessary, in the context of initial interest.

Seither-Preisler et al. (2007) depart from these perceptual studies by using intervals from the chromatic equal temperament scale. The use of musical intervals make the study more relevant to actual music perception. As we will see in Chapter 5, this expertise is based on a very unlikely artifact built into the stimuli, which puts into doubt any musical validity implied by the use of chromatic intervals.

On a structurally identical task to that of (Seither-Preisler et al., 2007), the results of the research reported in this thesis confirm that participants demonstrate performance differences that can be predicted by the number of years of formal musical training. The two-tone stimuli used here were not reported as musical by participants, although the intervals used between them were based on standard musical intervals of 2, 4, 5 ,7 and 9 semitones.

Unlike traditional psychometrics, where fine-grained, least noticeable difference thresholds and difference limens also suggest auditory expertise effects (Leipp, 1977; Rakowski, 1977; Meyer, 1978), the first replication in the upcoming research demonstrates expertise at intervals that are used in western musical contours. This expertise effect is surprising in that musical systems use intervals that greatly exceed the threshold of frequency discrimination where perceptual expertise between musicians and nonmusicians are shown to occur [around half a semitone (Meyer, 1978)]. Through the course of the experiments reported in this thesis, several potential confounds in the initial replication that may have been responsible for the effect of expertise are investigated and resolved.

1.11 Music

The present research uses the methods developed in the field of psychoacoustics where auditory stimuli are generated using additive synthesis, and participant responses are observed and analyzed. What such studies can tell us about music perception more broadly is often difficult to determine. Experimental economy allows us to reduce confounding factors and requires careful control of the properties of acoustic signals. Yet, the result of such control is that we must use stimuli that are remote from anything that a participant would identify as music. The focus is rather on the physical properties of the signal, and their perceptual counterpoints; as such, top down influences of active listening intentions, sensory motor feedback, as well as culture-specific factors of hearing/making sound are not considered in the design, nor analysed for in the data (see Bharucha, Curtis, & Paroo, 2006, for a review of the multiple implicit and conscious factors potentially at play in musical experience). Such reductive practices in psychoacoustics remain internally valid in that they are replicable in the laboratory setting. Yet, they do not extend to more holistic musical experience or performance (see Neuhoff, 2004, for a review of these reductive perils as well as possibilities for more ecological studies).

These laboratory methods have, nonetheless, revealed interesting perceptual limits, idiosyncrasies, and potential learning influences in human audition. Furthermore, these controlled environments can give us glimpses into the repercussions of excluding from analyses more ecological or embodied factors; in other words, the experiment can reveal in its outcomes the limits of its own controls. This revelation is what occurs in the course of the research reported in this thesis.

Retaining a perceptual context that participants' report as musical (i.e., with pitch contour, the criteria might be some recognition of melody or intended tone row) is of particular importance if we intend to elaborate on auditory effects that underpin pre-existing musical theory or even modify music theory through such explanations of auditory effects. Such research has been particularly fruitful as a perceptual counterpoint to the *post-hoc* extrapolation of musical rules from compositions in historical periods that constitute conventional music theory. Music theorists and perceptual scientists might sustain a beneficial tension between the operation of the auditory system and the intentions of composers. Yet, in contrast to these approaches that aim to elucidate why music is organized the way it is, Rosner (1988) describes perhaps the more common research program in psychology of music, where music is employed to reveal general principles of auditory perception. In this latter case, the subject of interest is what variables associated with music (in the case of the research reported here, musical experience or training) can tell us about audition in general.

1.12 Musical Elements in Reduced Stimuli

Scruton (1979) stresses that just as pitch is a psychological construct , the perception of an inter-note movement is also a subjective percept that has no material existence between the discrete tones (e.g., the 500 milliseconds of silence between the tone pairs in the research reported here). This capacity to hear inter-note motion is the basis for melodic perception. Having stated the disclaimer of ecological invalidity, the tone pairs are also not excluded from rudimentary auditory processes that are required for higher order music perception.

In fact, Hanslick's (1957) definition of music as any pattern of moving sound would not wholly disqualify such tone pairs, as these tones can indeed be heard to move from one to the other. That said, tone pairs might remain unsatisfactory as a pattern. Although the smallest perceptual unit for contour perception is indeed the tone pair, Warrier and Zatorre (2002) have shown that performance on pitch related tasks changes when that task is embedded in a broader tonal context, suggesting that tonal context interacts with pitch sensitivity, and we may see expertise effects that occur between tones, disappear within longer passages. Furthermore, Krumhansl and Shepard (1979) demonstrated that timbral interactions with pitch (an interaction that is observed in the research reported in this thesis) are much more apparent in two tone stimuli than in longer tonal diatonic sequences. The disturbance from interacting dimensions is attenuated in more musical contexts.

1.13 Pitting Timbre Against Pitch

The perceptual dimensions of pitch and timbre discussed earlier are not orthogonal perceptual dimensions. Seither-Preisler et al. (2007) used the effect of their interaction to amplify performance differences between groups of varying musical experience in a pitch contour detection task. They removed the F_0 from the tone pairs of their stimuli and introduced incongruous harmonic information between tones that further obscured the implied F_0 for many listeners. In the task of evaluating the direction of pitch shifts between these manipulated tones, Seither-Preisler et al. (2007) observed response patterns that suggested a strong effect of perceptual expertise correlated with musical training: musicians tracked the implied F_0 consistently, whereas nonmusicians were more likely to find the harmonics a rival perceptual focus, even though the smallest of the differences between the implied F_0 of the two test tones (2 semitones) is one even non-musicians usually have no difficulty discriminating when heard with the F_0 included. The manipulation of the interaction between harmonics and implied F_0 produced an expertise effect at musically relevant intervals. Seither-Preisler et al. (2007) inferred perceptual differences between those with and without musical training; it was suggested that musical training produced changes in the auditory system, and that these physiological differences were made manifest in musicians' consistent use of the implied F_0 in the task.

In the research reported here, we first attempted to replicate the expertise effect in the Seither-Preisler et al. (2007) task with our own materials and community of musicians and nonmusicians to demonstrate the robustness of the phenomenon. We then explore a series of variations on the task to isolate the source of the effect of expertise in their task.

The question of expertise is somewhat conflated with the question of where that effect exists, as a causal account seems to demand it. The dominant interpretation in the literature is that the expertise is in the actual hearing of the missing F_0 , as opposed to not hearing it at all. Seither- Preisler, for example, makes the claim that the effect manifests through the very perceptibility of the missing fundamental (Preisler, 1993; Seither-Preisler et al., 2007). Her methods, however, have not eliminated other possible effects that could account for musicians tracking the illusory fundamental more reliably. Preisler (1993) for example, gave participants a mistuned fundamental and asked them to tune it to the missing or illusory one. Regardless of what they were hearing, the skill of tuning was incidentally very important to the task despite its independence from the perceptibility of the missing frequency. It may, then, have been an expertise in tuning rather than a change in perception that led to the performance differences between musicians and nonmusicians.

More recently, Seither-Preisler et al. (2007) confirmed the expertise effect with a simplified task: participants were given two tones of differing pitches. The task was to specify whether they heard the second tone in any pair as rising or falling relative to the first. The tones were manipulated such that if the illusory F_0 fell then the overtone pattern rose. Participants capable of tracking the missing fundamentals must in some sense be able to hear through the conflicting movement of the overtones despite the fact that it is through those overtones that the F_0 is computed. Despite the apparent simplicity of the task, there is here again a possible confound; regardless of what is actually heard, the denotative precision of the terms "rising" and "falling" might vary as a function of musical training. The visual metaphor of height might only map precisely to pitch for a musician who has learned through ear training to use this vocabulary in a narrower sense. In a population without this experience, we might notice a broader use of the metaphor. It makes sense to say, for example, that the violin is a higher instrument than the cello, not only because it is played on average in a higher register, but also because of the qualities of its timbre; in this sense, our notion of the violin as higher, and cello lower are not reversed when we hear a duet between Mari Kimura playing subharmonics on her violin that fall deep into the cello's range, and a cellist playing in a register usually reserved for the violin. Is the expertise, then, merely the result of a honed metaphor in this case, because it seems that for some, both pitch and timbre will be subsumed in the metaphor of height? Indirect measures of pitch perception may be able to answer this question.

Chapter 2

Experiment 1a: Replication of the Seither-Preisler et al. (2007) AAT

Testing for sensitivity to the missing fundamental illusion is difficult. For example, a test that asks participants to report on whether a second tone was higher than a first in pairs that consist of overtone series at integer multiples of differing implied fundamentals contains an inherent confound; with most harmonic sounds, the relative heights of the harmonic frequencies shift congruently with that of the implied F_0, making any difference in perceptual focus impossible to observe. Even if participants were tracking only the frequencies of the harmonics, they would respond as if they were tracking the implied F_0. A frequently used alternative to eliciting responses of direction is to have participants tune a pure tone to the tone with the implied F_0, thereby demonstrating whether the participant matches the pure tone to the implied F_0 or some overtone component. This was, for example, the approach used in a previous study by Preisler (1993) to investigate the same expertise effect. Yet, this interactive task enlists skills that may be extraneous to the perception of the implied F_0; namely, the capacity to tune, and to know when you have a match (i.e., the consonant interval of the fifth, corresponding to the second overtone, might be mistaken for a match to a perceived F_0).

As a novel solution to this problem, Seither-Preisler et al. (2007) devised tone-pairs for which the frequency shift in the overtone series (using subsets of the harmonic overtones) was *inversely* associated with the F_0s they implied: when the implied F_0 of the second tone was *higher* than that of the first, the overtones used to imply that F_0 were *lower* in frequency than those of the first; similarly, when the implied F_0 of the second tone was *lower* than that of the first, the overtones used to imply that F_0 were *higher* in frequency than those of the first. Thinking in terms of voices or instruments, now, they paired the equivalent of a low bass note expressed as if it were played with an instrument that emphasised more of the high overtone frequencies (e.g., a violin) with that of a higher note, but played with an instrument that emphasised more of the overtones of the lower register frequencies (e.g., a viola), and vice versa. Indeed, although no real instruments were simulated, they paired various combinations (using different spectral series) of the same implied fundamentals with differing overtone series.

These stimuli provided the basis for what Seither-Preisler et al. (2007) referred to as the Auditory Ambiguity Test (AAT). In this way, if participants were merely tracking the frequencies of the overtones and not the implied F_0, their responses would now move in opposition to rather than in concert with the implied F_0, as they subsequently concluded for at least some of the pitch direction judgements of their nonmusicians, but not musicians. It is this result that is the focus of the current research. As such, it would seem prudent to replicate the results of Seither-Preisler et al. (2007) before attempting to explore the phenomenon with other manipulations. Accordingly, we produced tones and an AAT task similar to those used by Seither-Preisler et al. (2007), and applied them to our own population of musical experts, amateur musicians, and nonmusicians.

2.0.1 Method

Participants

Fifteen participants were recruited from the broader university community (i.e., undergraduate and graduate students, post-doctoral fellows, and faculty) at the University of Lethbridge. Through post-test interviews (but before performance was actually scored) participants were classified into three groups according to the extent of their musical experience. Those with over five years of formal musical training who were still involved regularly in musical performance were considered musicians ($n = 4$); those with one to five years of training were considered amateur musicians ($n = 4$); and finally, those who reported no musical training or participation were considered non-musicians ($n = 7$). The definition of musician here differs from that used in Seither-Preisler et al. (2007) where a more rigorous classical conservatory education and frequent practice was required. As subsequent results make clear, our definition also seems to select for participants with the appropriate differences in the levels of expertise.

Musical experience is indeed very difficult to quantify. These distinctions remain contentious as there are multiple kinds of musical experience that may influence the development of auditory perception. In future studies it might be preferable to make direct measures of musical ability using standardized tests as in Delogu, Lampis, and Belardinelli (2006); Slevc and Miyake (2006). Nonetheless, our criteria (namely, the first question in Table 2.1) are shown to replicate the expertise effect reported in Seither-Preisler et al. (2007).

For our purposes, anyone whose only experience was playing an instrument for 1-2 years in school to fill an elective, or as required by the curriculum (rather than for personal interest) was considered a nonmusician. More intensive training outside of school curriculum was considered more salient experience, and therefore a participant with 1-4 years in this case would be considered an amateur. The variability of musical curriculums in school and extra curricular training could not be controlled for and was accepted as a potential source of noise for the data.

Materials and Procedure

Tones were constructed to be structurally similar to those used in the AAT of Seither-Preisler et al. (2007) by summing for each tone sinusoidal waves corresponding to integer multiples of the frequency of the given F_0, and then normalising the intensity. The tones were digitised to 8-bit WAV files with a sampling frequency of 22050 (one-half CD-quality) and a duration of 500 msec. As in Seither-Preisler et al. (2007), these tones and tone pairs had the following characteristics:

1. the F_0s were not present;
2. overtones at integer multiples of the fundamental were selected so that the interval between the highest and lowest overtone was always an octave (a doubling of the lowest overtone frequency);
3. between the paired tones there was always a contrary motion in that, if the implied fundamentals rose, the overall shift of the overtones was down, and vice versa;
4. to vary the intensity of this conflict between the motion of the overtones and that of the implied F_0s, the tones were paired in three groups (A, B, and C) with different spectral profiles; in group A, the higher of the F_0s was implied by harmonics 2–4 (1 being the implied F_0) and the lower by harmonics 5–10. In group B, the higher of the F_0s was implied by harmonics 3–6 and the lower by harmonics 7–14. In group C the higher of the F_0s was implied by harmonics 4–8 and the lower by harmonics 9–18 (see Table 2.2). In using contiguous harmonics slight pitch shifts arising from the use of, for example, only odd, or only even harmonics were avoided (Meyer, 1978; Patterson, 1990).

The tones were selected such that the lowest implied F_0 was 100 Hz and the highest did not exceed 400 Hz. The interval between any given pair of implied F_0s was 2, 4, 5, 7, or

Table 2.1: Post-test questionnaire

How many years of music lessons have you had? From what age?
Did you do ear training exercises in your lessons or other wise?
Do you consider yourself musical?
Do you play a musical instrument? presently?
If so, do you practice every day? How many hours?
Have you participated in a choir or a musical group?
In this test did you sing or hum the tones?
If you hear a note in your vocal range, can you sing it?
Do you know where middle c is on the piano?
Do you ever tune a musical instrument?
If so, do you tune it using a tuning device like a pitch pipe or a tuning fork?
Do you ever tune a musical instrument to another musical instrument?
In general, do you find it easy to join in with other singers and sing in tune with them?
Do you often get songs stuck in your head?
At a live music show, have you ever noticed that an instrument or a singer sounds out of tune?
Describe your music listening habits (do you listen intently or use music as background for other activities).

9 semitones (half-steps on the Western musical scale). Therefore, the tones implying the lower F_0s in a given pair did not exceed 238 Hz to allow for an interval rising 9 semitones to suggest an implied F_0 that would not exceed 400 Hz. Within this range of implied F_0s of the lower tones, the following implied frequencies were used: 100 Hz, 123 Hz, 146 Hz, 169 Hz, 192 Hz, 215 Hz, and 238 Hz.

In Figure 2.1, an example tone pair from group A (see Table 2.2) is broken down into constituent frequencies. Where 1 is always the implied F_0 of 100 Hz, the first tone has sinusoidal overtones 5-10. The implied F_0 of the second tone is generated relative to the the implied F_0 of the first tone according to the chosen intervals of 2, 4 ,5 ,7 and 9 semitones. In the second tone of Figure 2.1, the implied F_0 rises 2 semitones relative to the first, which is 112.24 Hz. The arrows demonstrate the opposition between the contour of the implied F_0 and the movement of the overtones.

In sum, the combination of 3 spectral types $\times$ 5 semitone difference categories $\times$ 2 orders for each tone pair (ascending and descending) for each of the seven implied fundamental frequencies resulted in 210 unique trials. The order of these trials was randomized for each participant.

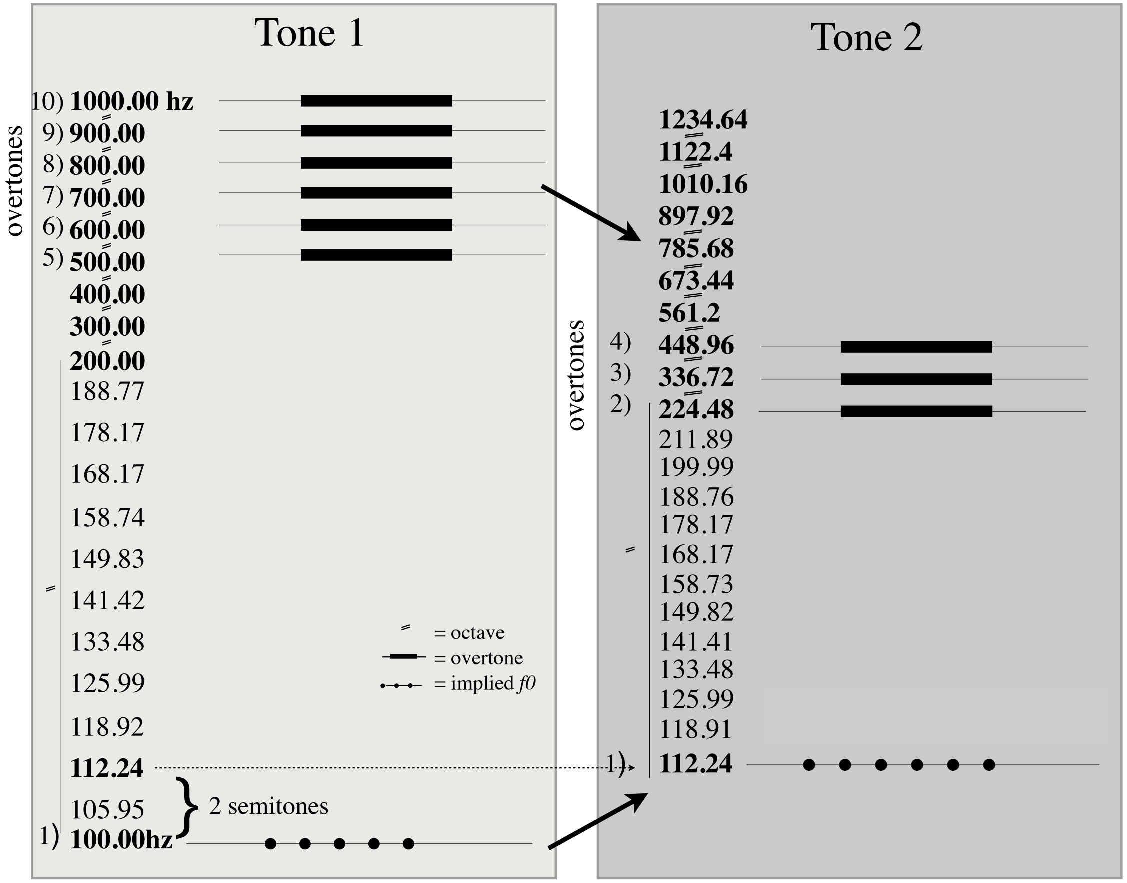


Figure 2.1: Example tone pair: 2 semitone rise from 100hz to 112.24 with overtones 5-10 and 2-4 respectively

Participants were seated at a computer and monitor equipped with headphones and a keyboard. Every participant was instructed on the use of the headphones (with the volume adjusted, if necessary, to accommodate individual hearing preferences) and the keyboard. Following presentation of the instructions on the computer screen, the testing trials began. For each trial, a large plus sign appeared in the center of the screen for 500 milliseconds before the tone pair played. Each tone played for 500 milliseconds, separated by 500 milliseconds. After the tone-pair sounded, the plus sign was replaced with a question mark to solicit the participant's two-alternative response choice. Participants were instructed to press the "?" key to indicate that they thought the tones rose ("went up"), or the "Z" key to indicate that they thought the tones had fallen ("went down"). Once the response was given, a "next trial" prompt along with the instructions: "press the space bar to move on to the next trial" were provided. As in Seither-Preisler et al. (2007), the instructions explicitly suggested that participants could sing or hum the tones if they thought it necessary.

2.0.2 Results and Discussion

The proportion of congruent responses (responses tracking implied fundamentals) were subjected to a 3 (expertise category) × 3 (spectral types) × 5 (semitone differences) mixed ANOVA, with subjects nested within expertise crossing spectral and semitone differences as the random variate. Figure 2.2 illustrates the mean congruent responses as a function of semitone difference for each of the three expertise groups.

As in Seither-Preisler et al. (2007), there was a significant overall effect of expertise on congruent responding $F(2,12) = 17.63, MSE = 0.133, p = .0003$ with musicians $(M = .94)$ evincing the highest performance followed by, respectively, amateurs $(M = .85)$, and nonmusicians $(M = .61)$. The mean difference between musicians and amateurs was not significant (minimum Fisher $LSD_{.05} = .145$), but both were significantly greater than

	Implied F_0	
Tone Pair Group	High F_0	Low F_0
A	2-4	5-10
B	3-6	7-14
C	4-8	9-18

Table 2.2: Grouped Harmonic Series That Imply the High and Low F_0s of the Tone Pairs in Experiments 1a and 1b.

the mean for nonmusicians (minimum Fisher $LSD_{.05} = .129$). Again, as in Seither-Preisler et al. (2007), there was a significant main effect of semitone differences $F(4,48) = 15.94$, $MSE = .018$, $p < .0001$; the larger semitone differences produced greater congruent responding than did the smaller ones: with a minimum Fisher $LSD_{.05} = .057$, all mean semitone difference comparisons were significantly different from one another except the two smallest (2 and 4 semitone differences) and the two largest (7 and 9 semitone differences). Indeed, as can be seen in Figure 2.2, nonmusician performance fell to chance (.50) at the smallest semitone intervals. The pairings of spectral profiles as summarised in Table 2.2 had no effect on participants' responses, $F(2,24) = 2.259, MSE = 0.017$, nor was there a significant interaction with either expertise, $[F(4,24) < 1, MSE = 0.017]$, semitone intervals $[F(8,96) < 1, MSE = 0.010]$, or their interaction, $F(16,96) < 1, MSE = 0.010$.

In contrast to Seither-Preisler et al. (2007), there was a significant interaction between expertise and semitone differences $[F(8,48) = 4.35, MSE = .018, p = .0005]$: both amateurs [linear trend: $F(1,12) = 59.65, MSE = 0.408, p = .0001$] and nonmusicians [linear trend: $F(1,24) = 62.13, MSE = 1.85, p = .0001$] tracked the F_0 more consistently as the differences between semitones increased. In contrast, the performance of musicians did not vary as a function semitone differences [linear trend: $F(1,12) < 1, MSE = 0.003, p = .5375$]: they were as accurate at the smallest iinterval (2 semitones) as they were at the largest (9 semitones). This result confirmed the conclusion that musical experience is correlated with response patterns in the AAT. It is from a similar result that Seither-Preisler et al. (2007) inferred hearing differences between groups. Replicating the effect was important to ensure that our stimuli produced at a minimum the same performance differences as that observed in Seither-Preisler et al. (2007) so that we could employ the stimuli confidently in other experiments.

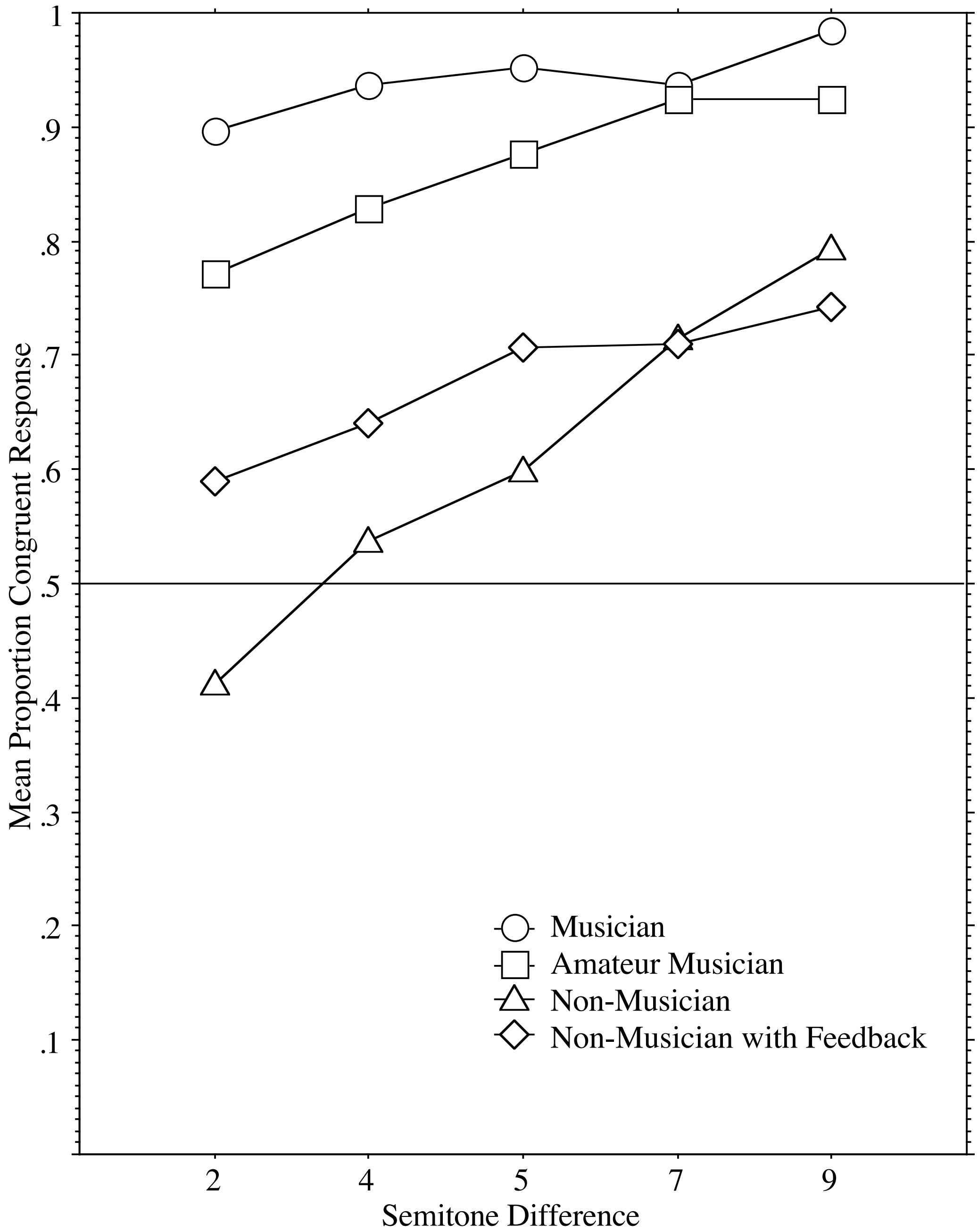


Figure 2.2: Mean proportion congruent responses as a function of semitone differences within the tone pairs and musical expertise in Experiments 1a and 1b.

2.1 Experiment 1b: The Role of Feedback in the AAT

One explanation for the different performance of nonmusicians, especially at the smaller semitone difference intervals, is that the metaphor of height may itself be ambiguous. In general, it has been shown that linguistic categories can influence perception (Stapel, 2007). Therefore, if musicians and nonmusicians are understanding the metaphor of height differently, it may be a source of performance differences. The words “up” and “down” applied to these ambiguous tones might be unclear to non-musicians, precisely due to these participants’ lack of experience with musical vocabulary in various contexts. Furthermore, height might be a metaphor for describing sound that crosses the exact domains being manipulated in the ATT such that “high” might refer, even if indirectly, to timbral qualities due to overtones as well as to pitch (Plack & Oxenham, 2005).

Here, the same orientational metaphor of height is used to map a dimension of timbre. In general, describing an instrument as being of a higher register seems to connote both a general range of pitches it can produce, but also its timbral qualities. We can also find specific examples where height is used to describe timbre [i.e., “an oboe has a higher spectral centroid than a french horn” (McAdamas, 2009, p. 73)]. This confusion may be a simple function of the natural positive correlation between the heights of overtones and that of their associated F_0. Another study by Boroditsky (2001) suggests that metaphors we inherit from our first language significantly shape the concepts we use to orient ourselves in the world. We might speculate that these kinds of linguistic effects observed across languages could be observed within as well; expertise, among other things, may be the result of acquiring a vocabulary, or refining an already existing one, with the subsequent effects on behaviour.

The problem that emerges in the AAT is that we are instructing the participants using a vocabulary that may only have a precise referent to those with musical training. Variable interpretations of the spatial metaphor of height used to describe pitch differences might be a source of performance gaps between those with and without musical training as explored in Experiment 1b. To rule out these semantic confounds, we tested a group of nonmusicians given feedback on each trial indicating how well their responses were tracking the implied F_0. If the lower congruent responding of non-musicians is a result of semantic confusion, the corrective feedback should provide a basis for disambiguating to what about the tones the terms applied, assuming it was available for them to respond to (see Goudbeek, Swingley, & Roel, 2009, for an example of trial by trial feedback facilitating sensitivity to acoustic dimensions).

2.1.1 Method

Participants

Thirteen nonmusicians (as assessed by the same post-test interview and criteria as in Experiment 1a) from the broader University of Lethbridge community served as participants.

Materials and Procedure

The materials and procedure were the same as in Experiment 1a, except that following the response, feedback in the form of “correct” or “incorrect” was displayed on the screen indicating whether the participant’s response tracked the change in pitch of the implied F_0s for that trial.

2.1.2 Results and Discussion

The proportions of congruent responses were subjected to a 3 (spectral types) $\times$ 5 (semitone differences) within-subjects ANOVA, with subjects crossing spectral type and semitone differences as the random variate. The results of this group are shown in Figure 2.2 as the group "nonmusician with feedback". As with the nonmusicians in Experiment 1a, performance for this feedback group increased as a function of increasing semitone differences of the tone pairs, $F(4,48) = 4.76, MSE = 0.031, p = .0026$. No other effects were significant. However, comparing nonmusician performance with and without feedback (i.e., the nonmusician groups from Experiment 1a and 1b) revealed that even though there was no overall difference in performance between the two nonmusician groups $[F(1,18) = 1.20, MSE = 0.257, p = .2882]$ there was a significant interaction with semitone differences, $F(4,72) = 3.826, MSE = 0.031, p = .0071$. Simple effects analyses of the interaction revealed that the feedback group performed significantly better than the no feedback group at the 2 semitone interval, $[F(1,18) = 5.064, MSE = 0.086, p = .037]$, but not at any of the larger intervals. It appears, then, that providing feedback was of some benefit, but only for the smallest of semitone differences. This small benefit of feedback in this task is consistent with the idea that at least some of the difficulties non-musicians have in tracking the implied F_0 is due to their less than clear understanding of how words like "up" and "down" map to judgements of pitch (i.e., F_0) as distinct from timbre (i.e., the frequencies of the overtone series). At the same time, it is clear that such semantic confusion does not provide any where near a full account of the expertise effect in the AAT.

Chapter 3

Indirect Measures of the Implied Fundamental

One possibility for the difference between musicians and nonmusicians in their ability to track changes in implied F_0 in Experiments 1a and 1b is that nonmusicians experience similar sensations/perceptions, but just do not respond explicitly to them as musicians do. To explore this possibility that responses, not percepts, differ, we sought a design that might reveal implicit influences of the implied fundamentals in the absence of the requirement that participants explicitly report on the pitch of the tones.

One of the ways to bypass the mediation of explicit influences of conscious decision making is the use of priming tasks in which reaction times are used to measure the presence of implicit sensitivity to the structure or characteristics of presented stimuli. In the following priming experiments, the relative height of AAT-like tones were used simply to warn the participant of an upcoming visual task. The warning tones' heights (of the implied F_0), however, were correlated with the location of the subsequent visual target. Implicit learning of this correlation might be observable in reaction times, which would suggest an implicit sensitivity to the implied F_0 without the influence of explicit decision processes.

In a series of priming experiments, we tested this hypothesis that an implied F_0 could implicitly cue spatial location of a target. If so, we would have evidence that, as with musicians more generally, nonmusicians can detect and respond appropriately to implied fundamentals in the absence of the requirement explicitly to label them.

3.1 Experiment 2a:

Priming for the implied F_0 with tones differing by an interval of two semitones

The performance difference between musicians and nonmusicians in the replication of Seither-Preisler et al. (2007) occurred for the most part at the smaller intervals. In order to isolate this condition where the strongest expertise difference emerged, only two tones were used that differed in frequency by 2 semitones (100 and 112.25 Hz). These two tones were no longer paired but used in isolation as primes for a spatial task on each trial.

3.1.1 Method

Participants

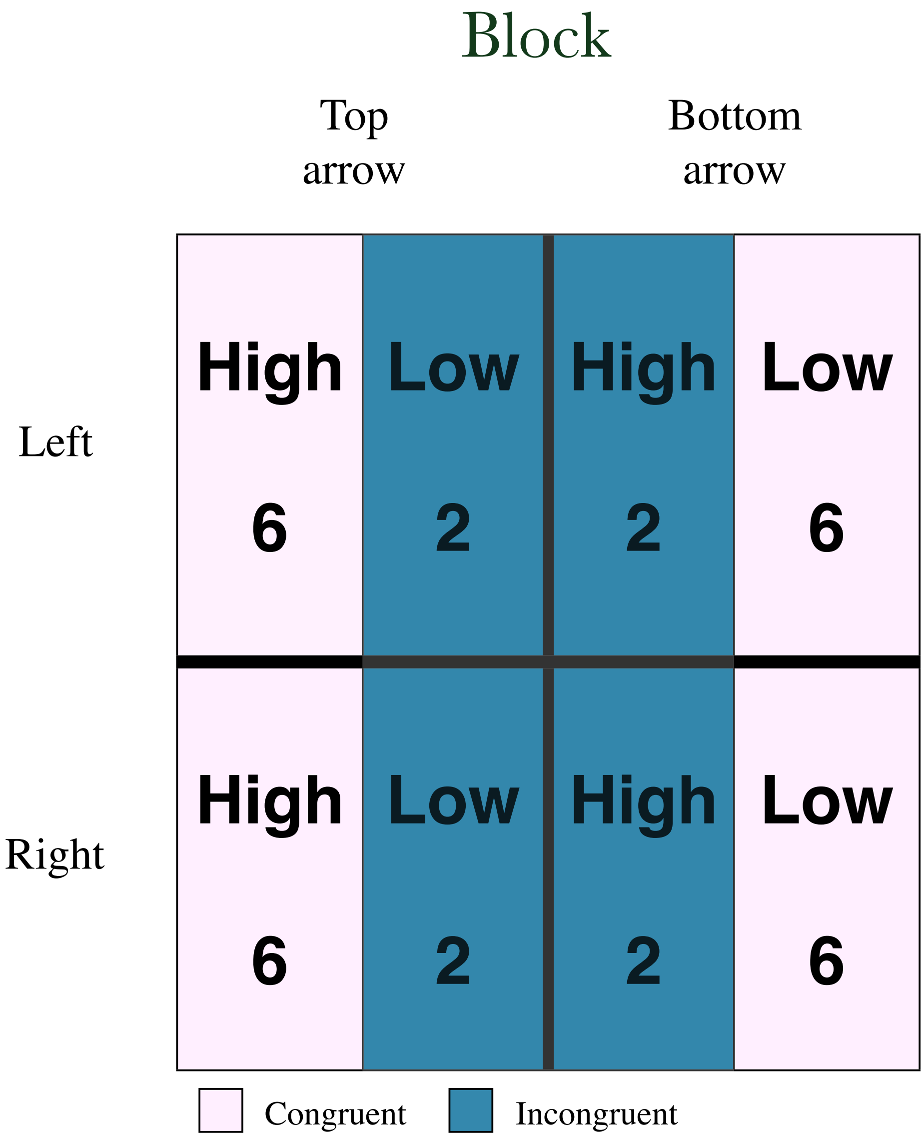
51 undergraduate psychology students participated in the experiment. Based on the same criteria as used in Experiments 1a and 1b, 18 were musicians, 18 were amateurs, and 15 were nonmusicians.

Materials and Procedure

Each tone varied in the overtones used to imply the missing F_0s. In parallel with the AAT, we used octaves within the harmonic series of each tone to imply the F_0; the tones, then, could have one of eight spectral (overtone) profiles: 2-4, 3-6, 4-8, 5-10, 6-12, 8-16, 9-18, 10-20 (where "1" is the implied F_0). The relative heights of the implied F_0s were positively correlated with the location of arrows that subsequently appeared on the screen. The tones were described in the instructions simply as warning tones for the subsequent appearance of visual stimuli. Following the tone with the higher implied F_0, 75% of the time an arrow would appear on the top of the screen 500 milliseconds after the tone. These were the congruent trials. On the remaining 25% of the trials—the incongruent trials—following the tone with the higher implied F_0, the arrow would appear on the bottom of the screen. In the same way, the tone with the lower implied F_0s predicted an arrow on the bottom of the screen 75% of the time with the same 25% remaining for incongruent trials. The participant's task, however, was unrelated to this correlation.

Each participant underwent 6 blocks of 32 trials each. Each block consisted of 8 trials that were incongruent (each representing one of the 8 spectral profiles) and 24 trials that were congruent (where each spectral profile was repeated three times) This structure of the blocks in terms of the proportion of congruent and incongruent trials is summarised in Figure 3.1

Participants were instructed to use keystrokes to report the direction the arrow pointed. The "?" key indicated an arrow pointing to the right; conversely, the "z" key was used to report arrows pointing left. The direction of the arrows was randomised and was not correlated with the implied F_0s. Participants were asked to respond as quickly and accurately as possible. It is important for subsequent developments in this research to note that, in order to test the possibility of auditory priming, participants were asked to respond as quickly as possible, something that in previous experiments was not required or encouraged. Reaction times (RTs) were recorded from the onset of the arrow on the screen until one of the two keys was pressed.



Block

	Top arrow		Bottom arrow	
Left	High 6	Low 2	High 2	Low 6
Right	High 6	Low 2	High 2	Low 6

☐ Congruent 🟦 Incongruent

Figure 3.1: Structure of one block of stimuli where High and Low refers to the relative height of the second tone relative to the first with the number indicating the quantity of such tone in a given condition.

3.1.2 Results

Errors

There was a significant reduction in errors in responding to the direction of the arrows over blocks, $F(5,240) = .001, MSE = .01, p =< .0001$, from a high of 4% on the first block to 1.1% on the last. This result indicates that participants improved at the task of reporting the direction of the arrows. There was no effect of congruency on errors, $F(1,48) = .002, MSE = 0.001, p = .5614$. Furthermore, as illustrated in Figure 3.2, there was no interaction between congruency and expertise, $F(2,48) = 0.002, MSE = .004, p = .5614$. No other effects were significant.

Error Corrected RTs

Reaction times on correct responses were subjected to a 8 (spectral series) × 2 (congruent/incongruent) × 6 (blocks) mixed ANOVA with subjects nested within expertise but crossing spectral series, congruency, and block as the random variate. There was an effect of block as participants improved at the task $F(5,240) = 2663.235, MSE = 12019.509, p = .0006$. However, here was no effect of congruency, $F(1,48) = 1536.462$, $MSE = 206.040$, $p = .7158$; participants were not significantly faster on congruent than on incongruent trials. As Figure 3.3 illustrates, there was also no effect of congruency by expertise on RTs, $F(2,48) = 1536.462, MSE = 590.410, p = .6830$; that is, even the performance of musicians, and not just that of nonmusicians, was unaffected by the implied F_0s of the tones.

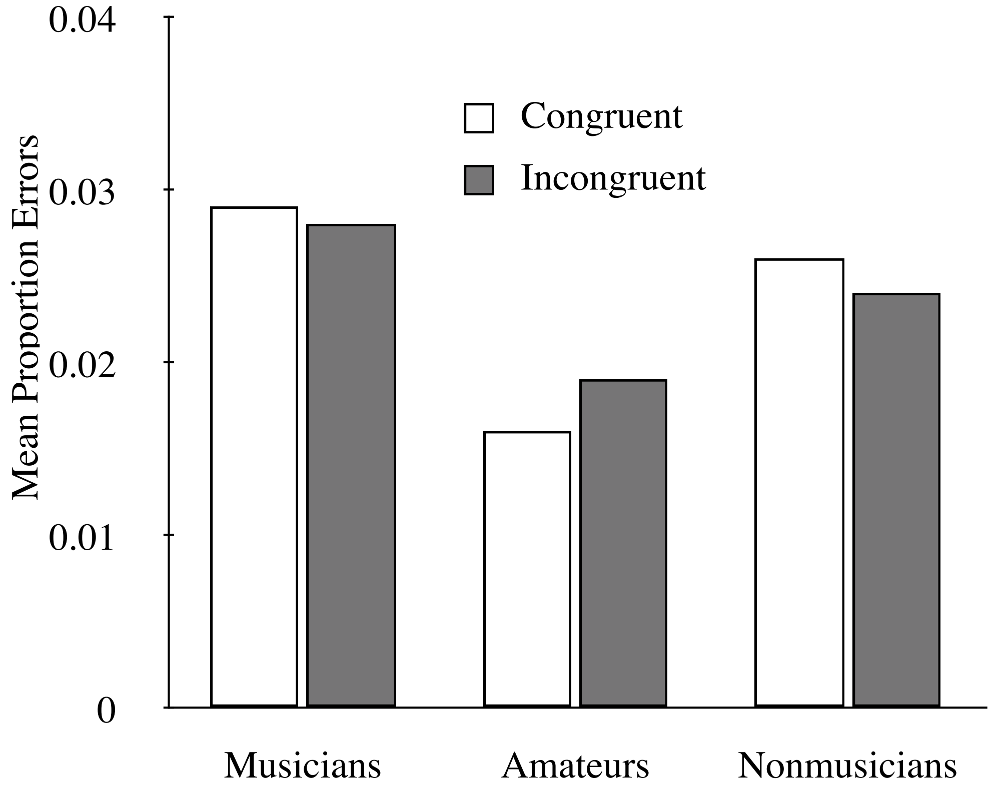


Figure 3.2: Mean proportion errors as a function of congruency and musical expertise in experiment 2a

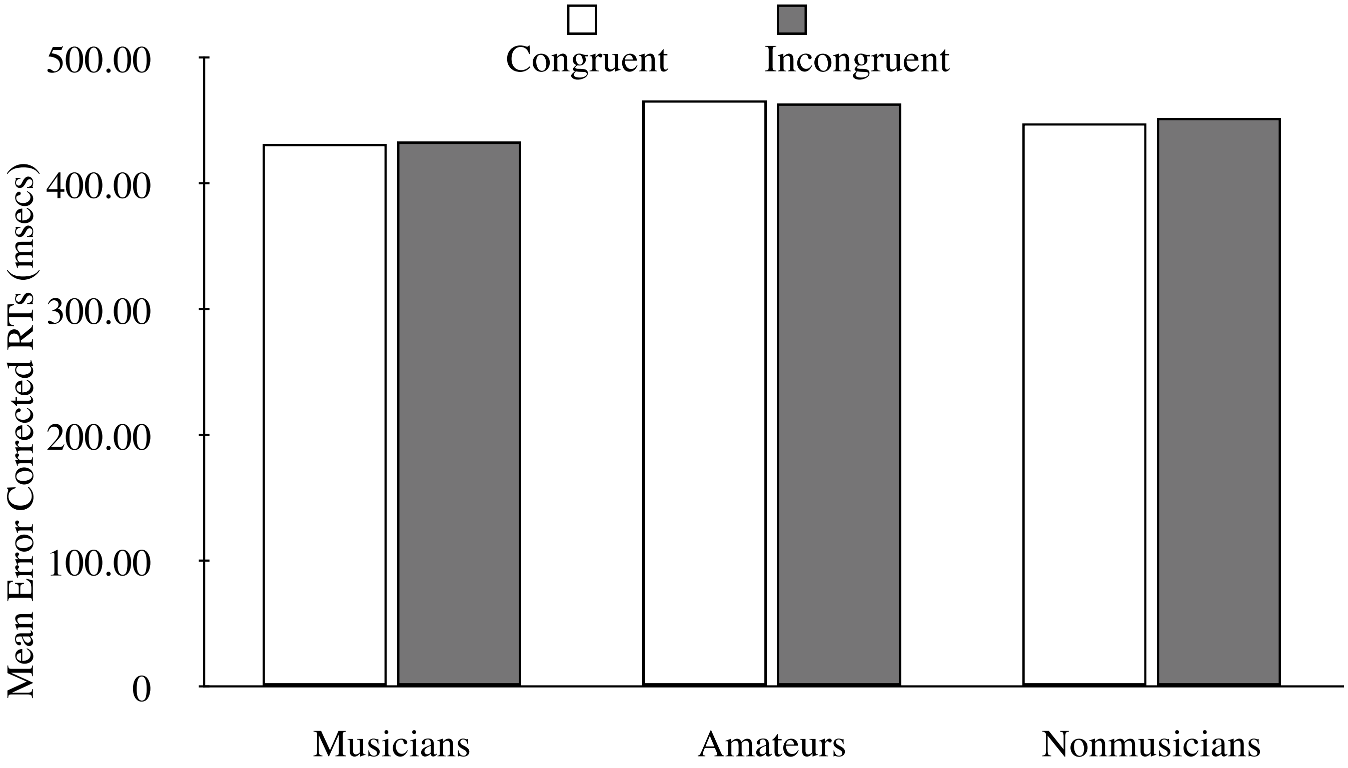


Figure 3.3: Mean Error Corrected RTs as a function of congruency and musical expertise in experiment 2a

3.2 Experiment 2b: Priming with the AAT stimuli

One possibility for the failure to find any effect of the congruency of the tones with the position of the arrows in the display even for musicians is that the implied fundamentals were only 2 semitones apart. Perhaps such a relatively small difference is not sufficient to prime different spatial locations. Furthermore, only a single tone was used on each trial. It is possible that using instead two tones moving in an implied direction, as with the AAT tones would produce superior priming of a spatial location. Accordingly, in the current experiment, we replaced the single tone primes from Experiment 2a with a subset of the AAT tone pairs from Experiments 1a and 1b. Intervals (rather than single tones) were now presented as primes that ranged from 2 to 9 semitones. In Experiment 1a we observed that across categories of expertise, all participants begin to track the F_0 at the higher intervals. This addition of a full set of AAT stimuli was done to see whether the priming of spatial location might emerge with intervals that induced more reliable F_0 tracking more generally.

3.2.1 Method

Participants

11 undergraduate psychology students participated in the experiment, 5 were musicians, and 6 nonmusicians based on the criteria as used in Experiment 1a.

Materials and Procedure

The tone pairs from the “A“ spectral series of Experiment 1a were used as primes, although only the first 4 fundamentals from that set were included. Each participant received two blocks of 160 trials each, obtained from the combination of 5 semitones × 4 fundamentals (100 Hz, 123 Hz, 146 Hz and 169 Hz) × 4 trial types (three of which were congruent, and one of which was incongruent) × 2 (within tone pair order: moving up or moving down). Order of the trials within the two blocks was random. Otherwise, the procedure was the same as in Experiment 2a.

3.2.2 Results

Errors

There were no significant effects of block, $F(1,9) = .001, MSE = .0002, p = .363$, congruency, $F(1,9) = .001, MSE = .001, p = .3721$, or congruency by expertise, $F(1,9) = .001, MSE = .0002, p = .6967$ on errors. There was also no effects of semitone $F(4,36) = .002, MSE = .001, p = .7552$.

Error corrected RTs

Reaction times to correct responses were subjected to a 5 (semitones) × 2 (congruent/ incongruent) × 2 (block) × 2 (musical expertise) mixed ANOVA with subjects nested within expertise but crossing semitones, congruency, and block as the random variate.

There was no effect of block, $F(1,9) = .001, MSE = .0002, p = .5615$. suggesting that participants were not varying their reaction times across blocks. There was no effect of congruency, $F(1,9) = .001, MSE = .001, p = .3721$, and no effect of congruency by expertise, $F(1,9) = .001, MSE = .0002, P = 0.6967$. There was no effect of semitones $F(4,36) = 388.030, MSE = 511.948, p = .2814$ indicating that priming effects were not observed at semitone differences where otherwise reliable tracking of the implied F_0 was observed across categories of expertise (as in experiment 1).

3.3 Experiment 2c: Explicit Responding

Experiment 2c was a replication of Experiment 2a with one critical change: participants were instructed explicitly on the correlation between the pitch of the two implied fundamentals and the subsequent location of the arrow, and that they could improve their speed of responding by exploiting that correlation. We revealed the structure of the experiment to the participants in order to see whether these tasks *could* produce differences in reaction times with overt use of the probability information.

3.3.1 Method

Participants

Eleven undergraduate psychology students participated in the experiment, 3 musicians, 5 amateurs and 3 nonmusicians. The criteria for these classifications were the same as in the previous experiments.

Materials and Procedure

The structure of this experiment was identical to Experiment 2a, with the one instructional difference: participants were told that 75 percent of the time the high or low tone, would be followed by the arrow on the top or bottom of the screen, respectively. They were then asked to use this information to predict where the arrow would be so as to perform the experiment more quickly.

As in Experiment 2a, each participant underwent 6 blocks of 32 trials each. The structure of the blocks in terms of the proportion of congruent and incongruent trials was the same as in Experiment 2a as summarised in Figure 3.1.

3.3.2 Results

Errors

Errors fell over blocks but not significantly, $F(5,40) = .007, MSE = .015, p = .0744$. There was no effect of congruency, $F(1,8) = .003, MSE = .716, p = .4220$, and no interaction between congruency and expertise, $F(2,8) = .003, MSE = .0004, P = .4220$.

Error Corrected RTs

There was no general increase in speed over block, $F(5,10) = 2.166, MSE = 5349.126, p = .0772$. There was also no effect of congruency, $F(1,2) = .549, MSE = 563.512, p = .4798$ and no interaction between expertise and congruency, $F(2,8) = 1026.051, MSE = 113.698, p = .8965$, on error-corrected RTs, similar to the results of Experiment 2a.

3.3.3 General Discussion

The inference of endogenous hearing differences between musicians and nonmusicians (as in Seither-Preisler et al., 2007) might exclude consideration of myriad other strategies, criteria and interpretations inherent in explicit responses—processes that may even determine the observed response patterns. The semantic issues associated with the metaphor (discussed in the preamble to Experiment 1b) are the kind of top-down influences that priming might control for and provide a bottom-up measure of a participant's sensitivity to the implied F_0. Yet, these three experiments showed no such effects even when participants were provided the correlates between tone height and target location.

Priming paradigms and the use of reaction times have now been used extensively in cognitive science; thus, we have a history of experimentation that suggests the contexts in which it does and does not reveal differences. Priming in the auditory domain has often shown no effects (Butchel & Butter, 1988) and in general, multimodal priming (i.e, an audio prime that is correlated with a visual task), has been very difficult to demonstrate. Nonetheless, Kawahara (2007) presents exceptional evidence that in certain conditions it is possible to get implicit spatial orienting from auditory cues.

One of the differences between Experiment 1a and the current priming experiments, is that, in the latter, the constraint of speeded responding had been added. The results of the experiments in the subsequent chapters suggest that such a constraint may preclude another important strategy that, indeed, may account for the original sensitivity of musicians to the implied fundamentals of the tones.

Chapter 4

A New Paradigm: Disambiguating Sources of Responding in the AAT

A principal difficulty with the Seither-Preisler et al. (2007) AAT is precisely the intentional ambiguity of the stimuli; by generating tones in which the overtone series are always inversely correlated with the shift in implied F_0, the confound referred to earlier in natural sounds (where overtones and F_0 move in parallel) is replaced by the reverse confound, making it difficult for the experimenter to tease out what participants were actually responding to. Following successful tracking of the implied F_0, for example, the AAT does not make it clear whether the participants were responding to the implied F_0, or the overtone frequencies responded to in reverse (a possibility that may account for the slight improvement at the two-semitone interval for nonmusicians given feedback in Experiment 1b). Not surprisingly, then, the ambiguity of the tones, presumably intended to challenge the perceptual focus of the listener, had the unintentional outcome of making the interpretation of the results ambiguous as well. Much of the Seither-Preisler et al. (2007) report explored a somewhat complex statistical approach in an attempt to mitigate this interpretation problem and confirm that, indeed, for the most part, nonmusicians were, if anything, tracking the frequencies of the overtones (especially for small semitones differences), while musicians were tracking the implied fundamentals. Accordingly, in this next experiment, we included conditions that allowed us directly to ascertain whether participants were tracking the implied F_0 or the overtone frequencies.

4.1 Experiment 3

The Seither-Preisler et al. (2007) AAT is somewhat complex for the purposes intended; on any given trial one of myriad initial tones could be presented followed by either a rising or falling tone of either high or low overtones implying their opposite missing F_0 frequencies. This variability in the initial tone is important for experiments designed to observe the ability of participants to detect (without the experience of a recurring reference tone) pitch differences (Burns E & Ward, 1978). For judgements of pitch direction, however, we may simplify the task by using a single reference tone, followed by tones that either rise or fall in implied F_0 by varying amounts from that reference tone.

To do so, we adapted an experimental paradigm of Rusconi, Kwan, Giordano, Umilta, and Butterworth (2006) that they used to demonstrate spatial response biases (i.e., responding faster to keys on the left [or down, such as the space key] for low tones, or keys on the right [or up, such as the 6 key] for high tones) that corresponded to the direction of real pitch shifts in F_0 and the corresponding overtone frequencies. Although the issue of response biases in reaction-time to real tone shifts is no doubt important, that was not our concern here, especially given the error-rate of non-musicians in Experiments 1a and 1b in which the fundamental frequency had been removed. There is no way that the corresponding high error-rate in the Rusconi et al. (2006) task with implied fundamentals could meaningfully be translated into a correct RT of any import. Despite that, we included each of the congruent and incongruent spatial response conditions of Rusconi et al. (2006), not so much to investigate spatial response biases, but to ensure that any such effects were counterbalanced over any other effects, and, hence, rendered independent.

The Rusconi et al. (2006) paradigm used a constant reference tone. The initial tone of all their tone pairs had a constant, real F_0 of 261.63 hz (middle C on the piano keyboard—key number 40). On each trial, the reference tone was followed by a second tone whose pitch either fell or rose relative to the reference tone by 2, 4, 6, or 8 semitones, and, hence, produced tones that also corresponded to notes on the piano keyboard. We used the same procedure except:

1. we used tones with the fundamental frequency removed, and
2. we used 3 different spectral (overtone) series to produce the implied F_0s of the second tones.

It was the use of these 3 different spectral series that provided a method for the disambiguation of the sources of responding.

4.1.1 Method

Participants

Eight musicians and 14 non-musicians (as assessed by the same post-test interview and criteria as in Experiment 1a) from the University of Lethbridge psychology undergraduates served in exchange for a partial course credit. Too few amateur musicians emerged from the post-test interviews to be included in the experiment.

Materials and Procedure

Tones were produced as in Experiment 1a, but at the semitone intervals used by Rusconi et al. (2006). The reference tone of an implied F_0 corresponding to middle C (key number 40: 261.63 hz) on the piano was produced using the octave overtone series of the fifth through to the tenth overtone (1308.13 hz to 2616.26 hz).

		Tones								
	Note	E2	F♯2	A♭2	B♭2	C3	D3	E3	F♯3	A♭3
	Piano Key	32	34	36	38	40	42	44	46	48
	F_0	164.81	185	207.65	233.08	261.63	293.66	329.63	369.99	415.3
Low	2	329.63	369.99	415.3	466.16		587.33	659.26	739.99	830.61
	3	494.44	554.99	622.96	699.25		880.99	988.88	1109.98	1245.91
	4	659.26	739.99	830.61	932.33		1174.66	1318.51	1479.98	1661.22
Medium	5	824.07	924.99	1038.26	1165.41	1308.13	1468.32	1648.14	1849.97	2076.52
	6	988.88	1109.98	1245.91	1398.49	1569.75	1761.99	1977.77	2219.97	2491.83
	7	1153.7	1294.98	1453.57	1631.57	1831.38	2055.65	2307.39	2589.96	2907.13
	8	1318.51	1479.98	1661.22	1864.66	2093	2349.32	2637.02	2959.96	3322.44
	9	1483.32	1664.97	1868.87	2097.74	2354.63	2642.98	2966.65	3329.95	3737.74
	10	1648.14	1849.97	2076.52	2330.82	2616.26	2936.65	3296.28	3699.94	4153.05
High	9	1483.32	1664.97	1868.87	2097.74		2642.98	2966.65	3329.95	3737.74
	10	1648.14	1849.97	2076.52	2330.82		2936.65	3296.28	3699.94	4153.05
	11	1812.95	2034.97	2284.18	2563.9		3230.31	3625.9	4069.94	4568.35
	12	1977.77	2219.97	2491.83	2796.98		3523.98	3955.53	4439.93	4983.66
	13	2142.58	2404.96	2699.48	3030.06		3817.64	4285.16	4809.93	5398.96
	14	2307.39	2589.96	2907.13	3263.15		4111.31	4614.79	5179.92	5814.27
	15	2472.21	2774.96	3114.79	3496.23		4404.97	4944.41	5549.92	6229.57
	16	2637.02	2959.96	3322.44	3729.31		4698.64	5274.04	5919.91	6644.88
	17	2801.83	3144.95	3530.09	3962.39		4992.3	5603.67	6289.91	7060.18
	18	2966.65	3329.95	3737.74	4195.47		5285.97	5933.3	6659.9	7475.48

Figure 4.1: Frequencies (Hz) of the Overtones Used to Imply the Missing F_0 of the Tones as a Function of Spectral Series (Low, Medium, and High) and Semitone (Note on the Piano Scale) for Experiment 3 and 4.

The structure of the 3 spectral series used to construct the second tones is shown in Figure 4.1, and labelled as Low, Medium, and High. The 8 tones constructed with the Low spectral series were each generated from the low octave of overtones 2 through 4. The effect of this construction is that each of the 8 tones, even those with implied F_0s greater than the reference tone have overtones either all of *lower* frequency than those used to generate the reference tone, or at least *lower* on average. Hence, for those tones with implied F_0s less than the reference tone, when played following the reference tone have the implied F_0s and the frequencies of the overtones moving in concert; in contrast, for those tones with implied F_0s greater than the reference tone, the implied F_0s and the frequencies of the overtones move in opposite directions, as in the AAT. The 8 tones constructed with the Medium spectral series were generated using the middle octave of overtones 5 through 10, the same as those used to generate the reference tone. Hence, for these tones, when played following the reference tone, the implied F_0s and the frequencies of the overtones always move in concert. Finally, the 8 tones constructed with the High spectral series were generated with the high octave of overtones 9 through 18. In this case, tones with implied F_0s greater than the reference tone have implied F_0s and absolute frequencies of the overtones that always move in concert. Conversely, those tones with implied F_0s less than the reference tone, when played following the reference tone, have implied F_0s and absolute (or averaged) frequencies of the overtones that always move in opposition, as in the AAT.

In sum, if participants track the implied fundamentals, then all three spectral series conditions should evince similar patterns of responding, and congruent responding should be high and relatively unaffected by semitone differences; if some participants are tracking the frequencies of the overtones instead, then this result will be revealed by opposite patterns of responding to the Low and High spectral series tones as a function of semitone differences.

Participants were tested as in Experiments 1a and 1b. As in Rusconi et al. (2006), test trials were divided into 4 blocks. The combination of 3 spectral series × 8 second tones resulted in 24 trials per block. The blocks differed in what response-mapping was used: whether the ‘z’ and ‘?’ keys or the space and ‘6’ keys on the computer keyboard were to be used for responding, and whether the ‘z’ (or space) or ‘?’ (or ‘6’) key was to be used to indicate an ‘up’ response.[1] The order of trials within blocks was randomised for each participant, as was the order of the 4 blocks. Furthermore, this set of 4 blocks was itself repeated 4 times, with a new randomisation of the order of the blocks (and trials within block) for each repetition, for a total of 384 trials.

Before each block, instructions appeared on the computer screen informing the participant of the mapping of keys to responses, and to place their index finger of their right hand on the ‘?’ or ‘6’ key, and the index finger of their left hand on the corresponding ‘z’ or space key appropriate for that block; this mapping remained depicted at the top of the screen for the block of trials. Participants initiated the block of trials by pressing the corresponding “up” key for that block. On each trial, a large plus-sign appeared on the screen, followed a second later by the reference tone for 1 second, followed by the test tone for 1 second. Participants were instructed to respond as quickly (including even during the playing of the second tone), but as accurately as possible, and that their responses were being timed. RTs were taken from the onset of the second tone, but were not used. Following the response, the screen was cleared for 1 second, then the next trial began.

4.1.2 Results and Discussion

The proportion of congruent responses (i.e., responses tracking the implied F_0) was collapsed over blocks and subjected to a 2 (expertise) × 3 (spectral series) × 8 (semitones)

---

[1] Rusconi et al. (2006) also crossed which hand was associated with the ‘6’ and (literally, in this case) the ‘?’ key, a manipulation we did not use here.

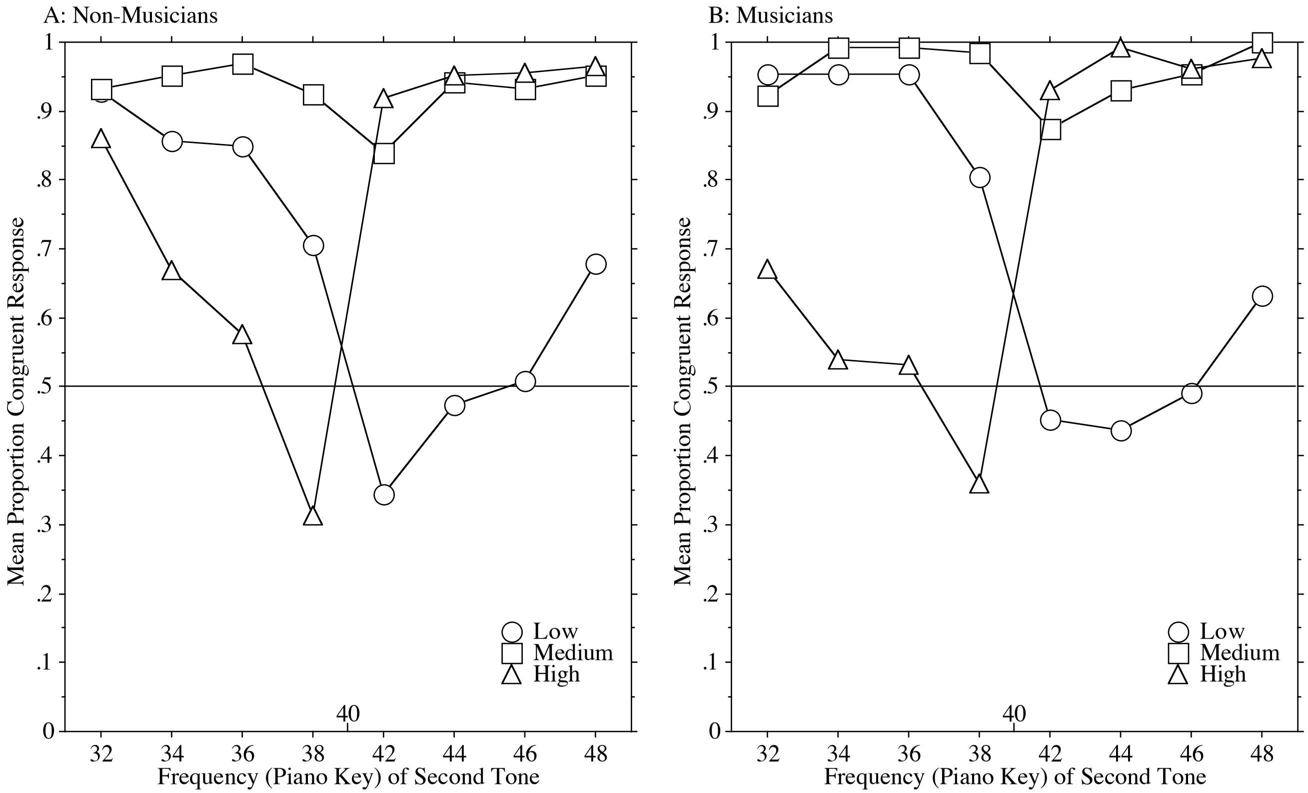


Figure 4.2: Mean proportion congruent responses as a function of spectral series (Low–circles, Medium–squares, and High–triangles) and semitone frequency (in piano key number) of the second tone relative to the reference tone (piano key 40) of the tone pairs for both Non-Musicians (subfigure A) and Musicians (subfigure B) in Experiment 3.

mixed ANOVA with subjects nested within expertise but crossing spectral series and semitones as the random variate. The results are shown in Figure 4.2, separately for the musicians and non-musicians.

Unlike in Experiment 1a, musicians ($M = .80$) were *not* significantly better than non-musicians ($M = .79$) at tracking the implied F_0s , $F(1,20) < 1, MSE = 0.018, p = .7323$. Nor did musical expertise interact with any of the other factors, either alone or in interaction, largest effect-size: $F(7,140) = 1.174, MSE = 0.02, p = .3218$. There was a significant effect of semitones, $F(7,140) = 13.858, MSE = 0.02, p < .0001$. As in Experiments 1a and 1b, on average, test tones with implied F_0s close to that of the first (or here, reference) tone resulted in poorer performance than those semitones more distant (for tones corresponding to piano keys 32 to 48, respectively the means were .89, .83, .81, .67, .72, .79, .8, .87). There was also a main effect of spectral series, $F(2,40) = 34.412, MSE = 2.798, p =< .0001$: performance was best with tones created with the Medium spectral series ($M = .94$), followed by the High ($M = .76$), and Low ($M = .68$) spectral series, with each mean significantly different from the others (Fisher $LSD_{.05} = .061$).

As is clearly evident in Figure 4.2, though, the major effect was the significant interaction between semitones and spectral series, $F(14,280) = 25.135, MSE = 0.038, p < .0001$. In general there was little effect of semitone differences with congruent tones (i.e., those with implied F_0s less than the reference tone for Low spectral tones, those with implied F_0s greater than the reference tone for High spectral tones, and all of the Medium spectral tones), and for which performance was quite good. Only those tones with incongruent movement of the implied fundamentals and the frequencies of the overtones (i.e., those Low spectral tones with implied F_0s greater than the reference tone, and those High spectral tones with implied F_0s less than the reference tone) resulted in a pattern of results similar to the AAT of Experiments 1a and 1b, and, as there, evinced a marked effect of semitone differences in the same pattern as in Experiments 1a and 1b for the nonmusicians.

As with the nonmusicians in Experiments 1a and 1b, participants in the current experiment evinced little ability to track the incongruent implied F_0s except for those tones most different in implied semitones. As noted, and unlike Experiment 1a, none of these effects varied significantly between the musicians and non-musicians: indeed, the musicians behaved as non-musicians in this task, showing no special ability to track the implied fundamentals.

Chapter 5

The Expert Voice

A category of possible explanations for the expertise effect in the AAT not explored by Seither-Preisler et al. (2007) is the local implications of musical experience on laboratory behaviour (i.e., what exactly musical training or being a musician promotes in behaviour associated with music); this behaviour may be directly related to the act of evaluating pitch. One possibility is the use of vocal strategies for dealing with pitch. For example, in the Seither-Preisler et al. (2007) task, comparative control would have been lost if their participants had been allowed to use a musical instrument through which they might try to reproduce the implied fundamental (by, for example, matching the overtone series) before responding. Musicians may, in fact, be more accomplished in the use of such a strategy that might not even occur to non-musicians, and therefore would have had an advantage that then would not be interpreted so much as a difference in *hearing*, but as a difference in *playing, performing,* or *acting*. The voice, albeit often characterised as distinct from other musical instruments (Cook, 1990), might be used in a similar way to disambiguate the stimuli.

Vocal strategies in hearing tasks are a recurring theme in auditory research. Thurlow (1963), for example, argued that vocal strategies can be used in the perception of the implied F_0. Thurlow and Hartman (1959) found that a vocal response (e.g., overt singing, subvocal humming, or even implicit singing) appeared to be necessary to mediate perception of the implied F_0. The source of response patterns to the implied F_0 in the Seither-Preisler et al. (2007) task, then, may be the very thing musicians are conventionally trained to do: the controlled production of sound through an instrument (i.e., in this case, the voice). The possibility that the source of the expertise effect in the Seither-Preisler et al. (2007) task may be that musicians and non-musicians act differently, rather than hear differently was explored in this next experiment.

Although there are any number of differences between Experiments 1a and 3, the most prominent is the demand in Experiment 3 (and 2a, 2b, and 2c) for speeded responding. If speeded responding is the reason musicians failed to show the expertise effect for the AAT stimuli, then there must be something that they do during the normal, non-speeded responding that nonmusicians normally do not do, or do substantially less well under the same conditions. The possibility explored here is that non-speeded responding provides for the opportunity to hum or sing the tones before responding, a suggestion provided by a non-musician who tracked the implied F_0 inordinately well for his expertise group; he reported humming the tones prior to responding in Experiment 3 (and, thereby, of necessity, ignoring the request for speeded trials). He reported that this vocal strategy would often reverse his initial response inclination, and even seemed to change his memory of the tones. Our post-test interviews in Experiment 1a also revealed one musician who acknowledged using the opportunity to sing the tones on each trial as a way of determining pitch direction. Furthermore, informal discussions with musicians more generally suggested that it is common knowledge, at least among those musicians required routinely to tune their instruments, such as string players, that humming or singing both the reference tone (e.g., a note played on a piano) and the tone from the string to be tuned is often helpful in determining pitch direction, especially as the difference between the two decreases.

5.1 Experiment 4: The F_0 Seeking Voice

To test the possibility that the elimination of the expertise effect in Experiments 2a, 2b, 2c, and 3, was a result of the requirement for speeded responding preventing a common vocal strategy for dealing with ambiguous tones among musicians, we repeated Experiment 3, but in place of the instruction for speeded trials, both musician and nonmusician participants were asked to hum or sing the two tones on each trial prior to making their response to that trial. The idea here was that if speeded responding prevented musicians from using a common strategy for them for dealing with ambiguous tones, and that strategy was to hum or sing the tones, then providing the requirement that all participants use that strategy, we might not only re-establish the musicians' ability to track the implied fundamentals, but possibly provide for the same ability in nonmusicians.

5.1.1 Method

Participants

Participants were 7 musicians and 7 non-musicians (according to the same post-test interview and criteria as used in Experiment 1a) recruited from the broader University of Lethbridge population of undergraduate, graduate, and post-doctoral students. Again, too few amateur musicians emerged from the post-test interviews to be included in the experiment.

Materials and Procedure

The materials and procedure for Experiment 4 were identical to those in Experiment 3, except that participants were instructed to respond as accurately as possible, and to hum or sing the two tones on each trial before responding.

5.1.2 Results and Discussion

The proportion of congruent responses were analysed as in Experiment 3. The results are summarised in Figure 5.1. In contrast to Experiment 3 (and 2a, 2b, and 2c), we observed

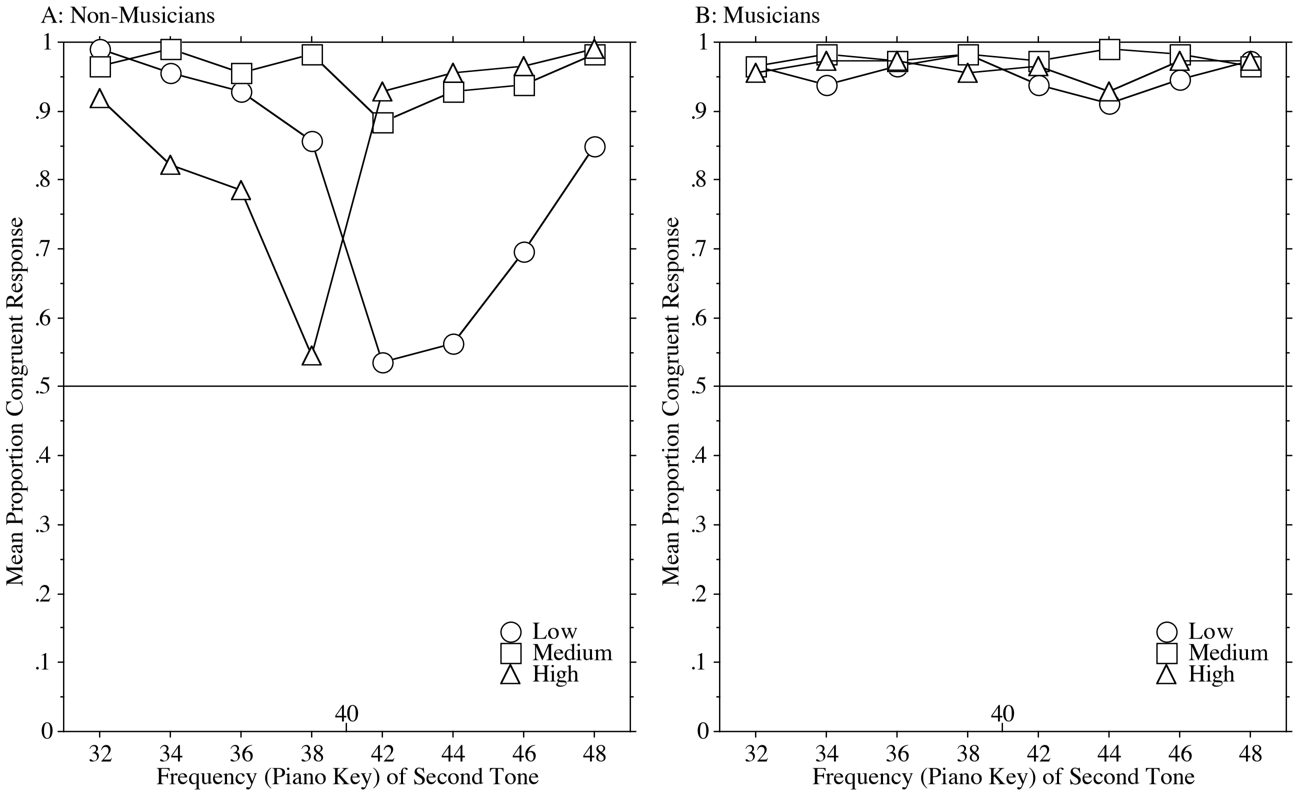


Figure 5.1: Mean proportion congruent responses as a function of spectral series (Low–circles, Medium–squares, and High–triangles) and semitone frequency (in piano key number) of the second tone relative to the reference tone (piano key 40) of the tone pairs for both Non-Musicians (subfigure A) and Musicians (subfigure B) in Experiment 4.

a main effect of expertise, $F(1,12) = 9.23, MSE = .077, p = .0103$. Unlike Experiment 3, but as with Experiment 1a and Seither-Preisler et al. (2007), musicians $(M = .96)$ tracked the implied fundamentals more consistently than did nonmusicians $(M = .87)$. As in Experiment 3, there was a main effect of spectral series, $F(2,24) = 8.67, MSE = .027, p = .0015$; the implied F_0s of Medium spectral series tones $(M = .96)$ were tracked more accurately than were those of the High $(M = .91)$ and Low $(M = .87)$ spectral series. There was also a main effect of semitones, $F(7,84) = 6.05, MSE = 0.009, p = .0001$: the more extreme the difference in semitones of the implied F_0s from the reference tone, the more the responses tracked the implied fundamental. Again, as in Experiment 3, there was a significant interaction between spectral series and semitones, $F(14,168) = 7.39, MSE = 0.011, p = .0001$.

However, as is evident in Figure 5.1, and unlike Experiment 3, the principal interaction effects involved expertise. The significant interactions of expertise and spectral series [$F(2,24) = 4.59, MSE = 0.027, p = .0204$], expertise and semitones [$F(14,168) = 7.39, MSE = 0.011, p = .0001$], and expertise with the interaction of spectral series and semitones [$F(14,168) = 6.03, MSE = 0.011, p = .0001$] all reflect the same result: musicians were unaffected by either spectral series, semitone differences from the reference tone, or the interaction of the two, evincing consistent and almost perfect tracking of the implied fundamentals in every combination, including the AAT tone-pair conditions [indeed, analysing for these effects on just the musicians revealed no significant effects: the maximum effect-size was for spectral series, $F(2,12) = 2.22, MSE = 0.0038, p = .1510$]. Non-musicians, on the other hand, were significantly affected by these factors and their interaction in much the same way as the participants in Experiment 3 [indeed, analysing for these effects on just the non-musicians revealed significant effects of both factors and their interaction: the minimum effect-size was for spectral series, $F(2,12) = 6.98, MSE = 0.0493, p = .0098$].

Relative to Experiment 3, musicians derived a clear and significant benefit from the requirement to vocalise the tones, to the point that we were able to re-establish the expertise effect of Experiment 1a and Seither-Preisler et al. (2007) for the AAT tone-pairs in the task. Was there any corresponding benefit of the requirement to vocalise the tones for nonmusicians? To answer that question, we compared the two nonmusician groups in Experiments 3 and 4. There was a significant increase in congruent responding for the nonmusicians in Experiment 4, $F(1,19) = 8.86, MSE = 0.081, p = .0077$, but none of the interactions between experiment and the remaining effects were significant [largest effect-size was $F(2,38) = 1.86, MSE = 0.057, p = .1672$]. In particular, despite the overall mean increase in tracking the fundamentals, and in direct contrast to the musicians, the nonmusicians in Experiment 4 were as influenced by the spectral series, the semitone differences, and their interaction as were the nonmusicians in Experiment 3.

The design of the priming experiments, and experiment 3, were motivated by our surprise at the expertise effects reported by Seither-Preisler et al. (2007) and replicated in our first experiment. These expertise effects were initially counterintuitive because the percept of an implied F_0 is commonly demonstrated with success to listeners of varying backgrounds. Furthermore, as already emphasized, the intervals used in (Seither-Preisler et al., 2007) exceed standard difference thresholds for contour sensitivity that might distinguish participants on the basis of varying musical training. With these factors in mind, we sought experimental designs that might reveal in nonmusicians a sensitivity to the implied F_0 through indirect tasks, sidestepping any complications associated with conscious, top down effects (i.e., presumed effects of vocabulary). In sum, we thought that through indirection, the nonmusicians might find direction out, when in fact, the musicians just got lost, leading to a serendipitous insight about the use of the voice in these tasks. In the concluding chapter, we propose several hypothesis on how such vocal feedback might function as well as possibilities for future research.

Chapter 6

Conclusion

Experiment 1a replicated the expertise effect in the Auditory Ambiguity Test of Seither-Preisler et al. (2007) using our own materials and populations of musicians and non-musicians, testifying to the robustness of the phenomenon. Using the same materials and task, we demonstrated in Experiment 1b that the source of the expertise effect was not due simply to semantic confusion on the part of non-musicians about the metaphor of pitch height. Experiments 2a, 2b, and 2c using a speeded, reaction-time priming task attempted to reveal a sensitivity to the implied fundamentals in nonmusicians using both single tones and the AAT stimuli, but found no priming effect, even for musicians. The subsequent manipulation of the AAT in Experiments 3 and 4 revealed that the expertise in this task was dependent on sufficient time following the presentation of the tone-pairs to allow musicians to do whatever it is that musicians do (and non-musicians typically do not) to track the pitch of the implied fundamentals of the tones. The question arises, then, what is it that the musicians are doing that requires this time when they perform distinctly from other expertise groups in the AAT? The most likely candidate, and the one these experiments suggest, is the very content of their expertise: reproducing tones.

In post-test interviews in Experiments 1a, 1b, and 3, few participants reported using their voices overtly in the tasks. One exception, as noted earlier, was the nonmusician who hummed the tones on each trial and tracked the implied F_0 quite well for his expertise group. His report that the humming would often reverse his initial interpretation of the pitch direction of the tones, and even alter his memory of them, is consistent with some promising lines of research and current theories of perception as performance or action (e.g., Clark, 1997, 2008; Hurley, 1998; Leman, 2008; Noë, 2004, 2009; Repp & Knoblich, 2007). Such feedback systems between active production and perception pose challenges to unidirectional models of hearing. Within such feedback models, responses cannot be reduced to mere outputs generated from stimulus inputs if they are in turn informing perceptions. Yet, psycho-acoustic research apparently does not dispense easily with these input-output models of perception; as such, Seither-Preisler et al.(2007) apparently did not consider the confound of vocal mediation in their pitch test of missing fundamentals. This omission is especially puzzling when we consider that they even suggested humming as a strategy for their participants, without then controlling for it as a variable and possible source of response differences. Such lacunae in the analyses of data might be the inevitable outcome of a pervasive computational analogy where response patterns are assumed *a priori* to have a monotonic and sequential relationship to the physical properties of sound (input) entering the auditory system (CPU) in order to achieve some response (output). This kind of assumption leaves little room for sensorimotor feedback proposed over a century ago by Dewey (1896), and to which our findings lend support.

This systemic bias can obscure recognition of embodied sources of perception inherent in many psycho-acoustic tasks (see Blesser & Salter, 2007, for a discussion of the difficulties with subjective judgements in psychoacoustics). These exogenous sources of variance inevitably alter what the word “hearing” might refer to.

Hypotheses of both covert and overt vocal mediation of pitch judgements emerge from these findings. An overt mediation hypothesis consistent with the early work of Thurlow (1963) would state that the musicians in our research know how to use their voices and know when they have a voice-tone match. Such a strategy would produce the overtones simultaneously by matching implied F_0s and thereby disambiguate the pitch direction: Even if the musicians were simply trying to reproduce the frequencies of the overtone series of the two tones on any given trial, they would necessarily have to produce the implied fundamentals of the tones to do so. This strategy is also consistent with studies that demonstrate generalized attuning abilities (e.g., Heylen, Moelants, & Leman, 2006). Trained singers and untrained participants with talented singing voices are more adept at pitch mapping than are those with no training and who report no singing talent (e.g., Watts, Moore, & McCaghren, 2005; Watts, Murphy, & Barnes-Burroughs, 2003). Such demonstrations of general abilities as well as expertise might lead us to explore the hypothesis of overt pitch mapping in the context of the AAT. Yet, if in our Experiment 1a, the vocal mediation of the successful tracking of our musicians was overt, it was at least not reported often as such; as mentioned previously, in our post-test interviews, few of the musicians reported the use of overt vocal strategies.

Covert embodied responses to hearing tones might also occur, and may mediate explicit assessments of tones; if so, we might expect the possibility of the embodied response as somewhat independent from the explicit judgement about the tones. Loui, Guenther, Mathys, and Schlaug (2008) demonstrated such independence when they found that tone-deaf participants often demonstrate a mismatch between the significantly above-chance accuracy of the pitch direction of their humming and the chance accuracy of their subsequent judgement when evaluating the direction of F_0 pitch shifts in tone pairs. This mismatch between active production and judgement indicates a kind of embodied knowledge to which these participants seem not to have access, do not use, or do not know how to use.

A related mismatch was demonstrated using electroencephalography (EEG); when presented with pitch anomalies in melodic passages, the patterns of EEG activity correlated with normal pitch detection did not differ significantly between congenital amusics and controls (Peretz, Brattico, Järvenpää, & Tervaniemi, 2009). It seems the expression of amusia is not to be found in neural substrates associated with this measure. These EEG results provide additional evidence that overt responses often are a poor measure of the underlying physiological sensitivity in pitch-related tasks.

Other results might challenge any overstatement of this decoupling of action and perception. Dalla Bella, Giguère, and Peretz (2009) confirmed consistent impairment among amusics in both production and perception of pitch changes. Yet, they also report exceptional cases where amusics with impaired melodic pitch discrimination produce normal melodic pitch intervals when singing. An embodied sensitivity to pitch changes may, therefore, not be ubiquitous among amusics, yet its presence, even in a subset of a population usually considered to have no access to normal pitch detection, suggests that it may be playing a much larger role where, as in the current experiments, there is no self-reported amusia. The results of such studies comprise growing evidence that embodied cognition can happen independently from the accurate application of spatial metaphors such as height in pitch perception. The association between action and perception in this context may be extended to a spectrum along which we might attain a good predictor of musicality; the degree to which the vocal action and the judgement of pitch direction are coupled may vary, with musicians possibly evincing the strongest association. Thus, it is quite possible that the nonmusicians in Experiments 1a, 3, and 4 are capable of tracking the pitch direction of the tone-pairs in performance (e.g., humming or singing), but merely fail to incorporate that performance accuracy into their subsequent judgements of pitch direction.

Our instruction to hum or to sing in Experiment 4 encouraged the strategy of vocal mediation across categories of expertise. Nonetheless, it was only the musicians who were consistently sensitive to how such humming or singing (or their necessary production correlates) disambiguate the pitch direction of the tones. One possibility here is that musicians really are just more accurate in performing pitch shifts than non-musicians are, and exploit that talent when given sufficient time to do so. We are currently exploring that possibility by recording the vocal reproduction of the tones of both musicians and non-musicians in the AAT.

Another possibility is that is it not so much differences in pitch reproduction between musicians and non-musicians *per se* that matters, but, as suggested by the aforementioned results of Loui et al. (2008), it is that musicians but not nonmusicians know, implicitly or explicitly, both how and what to use from such vocal productions to assess pitch direction. One function of musical training, for example, may be to bring these associations into the repertoire of tools of the practising musician. If so, we should be able to block that strategy in musicians by requiring some other, unrelated vocal production (e.g., the requirement for a monotonic "la, la, la, ..." or "January, January, ...") on each trial (Brooks, 1968).

It is quite possible, of course, that few musicians ever hum or sing as a strategy to assess pitch direction, as they do not need explicitly to do so. That is, it may be something inherent in humming or singing, but not humming or singing themselves, that is the true source of the musicians' ability to track changes in pitch height in the AAT (or otherwise), given, as in Experiments 1a and 4, sufficient time to do so. But, whatever it is it has to be at least compatible with such humming or singing, otherwise the requirements of Experiment 4 should have eliminated it, as did the requirement in Experiment 3 for speeded responding. Thus, subtle throat movements or even attempts to imagine the tone pairs *and sensitivity to the consequences of those actions* (Thurlow, 1963; Thurlow & Hartman, 1959) may be sufficient to explain the ability of musicians to track the missing fundamental in the AAT.

Sensorimotor imagery (a mental process that could potentially take as much time as the action) might become coupled with the imagery of timbral shifts that such vocalized actions necessitate. Motor imagery and timbral imagery may be synonymous here in a cognitive sense. Such a line of research might elaborate on Crowder's (1989) fascinating result where imagery of timbral qualities influences the acuity of pitch judgments. This modification of sounds through mental image might be interpreted as a kind of priming. Yet, the present work suggests that attributes of the stimuli, reported from memory, might also be modified *post hoc* when mental images are produced.

Pitch perception is integral to music as well as language perception. Although little has been done to elucidate the role of vocal responses in music perception, the parallel domain of speech perception has accumulated much evidence sinceLiberman's (1957) seminal work on a motor theory of speech perception. In this work we see that perception may be coupled strongly with the gestures implicit in their production (see Galantucci, Fowler, & Turvey, 2006, for a review of the motor theory of speech perception). Liberman (1957) asks, “when articulation and sound wave go in their separate ways, which way does perception go”. The question generates potential research directions here. For example, in the AAT, we can now explore response patterns when the tones from the AAT are heard during an imposed incongruent vocal gesture and other manipulations. The aforementioned participant who reported vocalizations as reversing initial response inclinations suggests that indeed perception can follow articulation rather than auditory event. We are also currently exploring placing the same AAT tones in seemingly (to the participants) vowel-speech contexts to see whether that would exploit the latent speech expertise of nonmusicians in the AAT.

Chapter 7

A circuitous walk through the laboratory of ventriloquy at the Banff Centre for the Arts

7.1 Analogous Fields: Arts and Science

In addition to the research completed in professor John R. Vokey's micro-cognition laboratory at the University of Lethbridge, a six week visual arts residency was completed at the Banff Centre for the Arts with Denton Fredrickson. The theme of the residency was Analogous Fields: Arts and Science. Art Critic Saul Ostrow and Artist/biologist Charles Tucker invited collaborative groups to propose works to explore how methodologies of art and science are not simply complementary, but analogous in the modes through which they generate knowledge. Such analogies are commonplace. Leonardo da Vinci is perhaps too often enlisted as the mascot for such analogies because of his activities in both domains. Yet, his own work may also illustrate how general descriptions of these methodologies puts them in elegant opposition rather than in any way analogous. For example, it is the assumed uniqueness of a result that defines his paintings in the Louvre, and accounts for the onerous security system that protects them, whereas the high ideal for experimental science is it's transparent replicability in any laboratory.

Nonetheless there may be methods that are common to certain art practices and science such as the process of experimental/creative control—that at first glance, paradoxical activity of removing elements of the world as a means of discovery. This notion of removal became the theme of the initial proposal.

7.2 Project Proposal

The proposed project is to produce an evolving interactive electroacoustic sound installation/instrument. As a departure we will investigate the notion of the instrument(al) in science and art, with a focus on strategies of removal. From this idea, an analogy is explored between seemingly disparate experiences in art practice and a psychoacoustics laboratory.

In an experiment on a reported perceptual expertise effect among musicians, Granzow and Vokey (2009) discovered that the ability to track the missing fundamental in the laboratory is dependent on the listeners' active voicing (whether subaudibly or aloud). The emerging questions are not so much about whether the ability to detect sound is revealed through some instilled or genetically determined neural architecture, but, rather, the role of embodied processes in such abilities.

The notion of embodiment and action as inseparable from perceptual processes is receiving increasing attention in cognitive science and robotics. Nonetheless, the assumption of perception as a unidirectional response to stimuli persists in psychoacoustic research; the mode of inquiry itself assumes a tight nexus of causality between human responses and the auditory stimuli. For purposes of analyses, these stimuli are often simplified using computer synthesis to reduce possible sources of variability in responses. Researchers in the laboratory use instruments of removal that control acoustic variables in any given experiment.

Various strategies on the part of historicized aesthetic movements (concept-based and minimalist composition, for example) could be said to engage in analogous kinds of removal. Removing frequencies from conventional sound envelopes by placing bolts in the strings of the piano, or to leave the piano completely silent (to borrow from the experiments of John Cage), is to use the object to remove its presumed and historically contingent sound. Sub-audible humming of participants in a psychoacoustic laboratory could be compared to the coughing and murmuring at a performance of Cage's 4'33; in both cases, instruments of removal (silent piano, lab software) become mediated with an instrument of restoration: The voice. The voice, in these contexts, can be seen as restoring experiential elements into the world and reinstating a kind of active texture.

The possibility of instruments involving the seeming paradox of removal as an agent of restoration through embodied responses of active participants, will be the foundation of our exploration of analogies between qualitative and quantitative practice. Our work will emerge from a dialogue between what has often been deemed a scientific, reductionist approach, and various historical aesthetic approaches (such as process-oriented sculpture/installation, minimalism, and/or conceptual art).

7.3 Removing the Air: Robert Boyle's Pump

In Joseph Derby's painting, "An Experiment of a Bird in the Air Pump", a little white bird, deprived of the spring of the air, expires in Robert Boyle's now iconic vacuum chamber. Boyle's demonstrations appealed at once to the veracity of eye-witness (although observers more empathic than curious are depicted as looking away), while emerging from a philosophical tradition that increasingly held the senses suspect of deception. The macabre outcome of his performance was to suggest robust conclusions about the nature of air, conclusions uncorrupted by introspection or the narrating voice; it was the fulfilment of "Nullius in Verba" (on nobody's word), the motto of the Royal Society that Boyle was so influential in shaping. It seems that in the ideal of this society the voice was disqualified from inquiry. Apposite then that the painter would depict Boyle as eliciting no outward signs of verbal accompaniment. His lips are closed. He seems to look out of the painting with portentous recognition. What can we assign to that look? Did the experimental criteria of replicability apply to the initial wonder as well, each demonstration as remarkable as the

Figure 7.1: Joseph Wright of Derby: An experiment of a bird in the air pump.

first? Or is he experiencing some vestige of the very introspection such empirical instruments were meant to replace? Or just maybe he is troubled by a glimpse into the future, his glass globe briefly exchanging data for clairvoyance where he spies a developed bow valley, and a pump that not only contradicts his law of soundless vacuums, but might even cajole him, berate him and, like Charlie Mcarthy [the now archetypal ventriloquy dummy from 1940's and 50's American radio and television], draw attention to the surreptitious throat movements of his interlocutor.

7.4 How it came to pass that a ventriloquial dummy was placed in the pump

Excerpt from narrative written on the wall in the laboratory at the Banff Centre:

> "I acted (I was acting anyway) as if the abduction of my dummy right from over my arm did not bother me. As they fled, the masked acousticians shot back armed looks to see, to their surprise, that I did not appear at all anxious to pursue them. The jaw of my dummy flapped loquaciously as they jostled it, and so I threw my voice into the fray in such a way as to reveal why I could only be thrilled with the way things were unfolding.
>
> "Finally shucked your arm of me I see. I'll curl my snickersnee (I would have said dagger here, but one of the men tripped, and the hinge I had carefully crafted many years hence bobbed tri-syllabically) around the throat of your thumb again, be sure of it."
>
> Having established in the abductors this belief that they had in fact delivered me from captivity, their vigilance subsided. It was at that point that I became

> a shadow, moving as they moved, pulled along through the alleyways and corridors, throwing my voice in supplication, and more importantly throwing it in concert with the terrain that rattled my dummy's jaw, (the way the texture on a cylinder might vibrate a stylus), to sustain from the grottos of their circuitous wake the wonder that could only inspire such theft."

7.5 The Voice

In psychoacoustics, the relationship is sought between acoustic properties and perceptual properties and thus vocal mediation can have significant effects on response patterns in the laboratory. Just as a ventriloquist deftly provokes the misattribution of voice to the animated dummy, so musicians may restore anticipated congruency (i.e., parallel contours between overtone and implied F_0) to the signal through unconscious vocal mediation. The vaccuum has become emblematic for the scientific method where the world is not observed as it is, but controlled through the process of removal of elements of the world itself to infer causality (Schapin & Schaffer, 1989). If removing the air removes the aria, then sound must depend as a medium on what was removed. Applying such simple logic of causality to perceptual studies without considering the creative capabilities (vocal interpolation) of the participant can be the source of misleading conclusions.

The analogy developed in this particular part of the installation hinged on active listening. In music, active listening is to attend acutely to the way the sound is organised. On a smaller scale it can include an almost conscious attempt to perceive component parts of a sound that under passive listening might seem singular (as in the case of Mersenne, 1636). Take a struck bell. Passive listening would be to hear the sound as a whole, perhaps even hear a semantic level: order is up, shift is over, grease fire, propinquintous bear. Slightly

Figure 7.2: Ventriloquy dummy in Boyle’s air pump

maladapted to these cues, active listening would be to do such things as count how many harmonics can be heard as individual tones in the sound. Active listening is extended here to actions themselves. To listen actively is potentially to employ embodied strategies to feed back into perceptual focus. In the context of the AAT where the task solicits a forced choice response to a stimulus that contains rival perceptual foci, these embodied reactions become integral to the performance differences between those with and without musical training.

During the residency, the missing fundamental illusion was recast as the ventriloquial dummy-tone, where what is missing from the tone is filled in with a projection of the voice. The careful control of the physical properties of stimuli correspond to a logic of inferential statistics; but participants, in turn, act to restore an anticipated congruency to the stimulus.

7.6 Rhythmicon

Gibbs (2005) observes that in indo-european languages “to see” and “to know” often become synonymous. Many of our metaphors for describing the other sensory domains come from vision. Acoustic signals are often represented graphically as a set of discrete component parts or frequencies. This spectral representation of sound emerged from Fourier’s insight that any wave form could be seen as the sum of various sine waves. These components are often represented as vermiform lines between two axes: frequency and time. This visualisation of sound may contribute to the presumption that we can simply assign various auditory phenomena to sonic counterparts, like monsters to their makers. Yet we rarely experience sound in such a particulate way, and the perceptual interactions of component parts of a sound elude simple isomorphisms between physical properties and perceptual experience. There is a tension, then, between our visual representations/metaphors and our more protean auditory percepts. In response to these considerations, we built a version

Figure 7.3: Distributed rhythmicon

of Henry Cowell's rhythmicon that was strung across a structure in the form of a three-dimensional graph. The axes, like the jaws of a trap, were set to ensnare the spherical wave forms of their own sounding. But the sound got away. This tension between visual representations/understandings of sound and the sound itself can cycle in this kind of research, the phenomenal experience giving way to our visual de-constructions and back again to our renewed and perhaps altered listening. What has been the influence of this tension on the way we study sound, the way we listen to it, even the instruments we might build?

7.7 The Daxophone

In response to these questions, we also made a set of Hans Reichel's Daxophones that were constructed with a feed-forward process that went from a visualisation of sound energy to the use of those visualisations in the generation of unique timbres. To make the daxophones, we held a cherry plank at the 1st node, knocked it, recorded it, and generated a spectrogram from the recorded sound. We then removed just enough of wood from the plank to make the first daxophone (see Figure 7.4). The first spectrogram from the knocked plank was used like a stencil to determine shape of the first daxophone tongue that in turn determined the timbre of the instrument. Knocking the retreating plank again, a new tone at a different pitch and with a different spectrogram would emerge. This spectrogram was used to produce the next iteration of daxophone tongue, and so on, and so on.

In the studio, the voice-like sounds of the daxophones were amplified behind the dummy in the pump. The work is still in progress. The apparent voice of the dummy has its source in any proximal speech that coincides with the operation of the crank. Yet we await the concomitant fricative with the friction of horse hair and rosin on the wooden tongue of the daxophone. During our open studio, it happened once, and the pump operator leapt back.

With the help of Chris Chafe at Stanford and Doug Van Nort at McGill, we are now

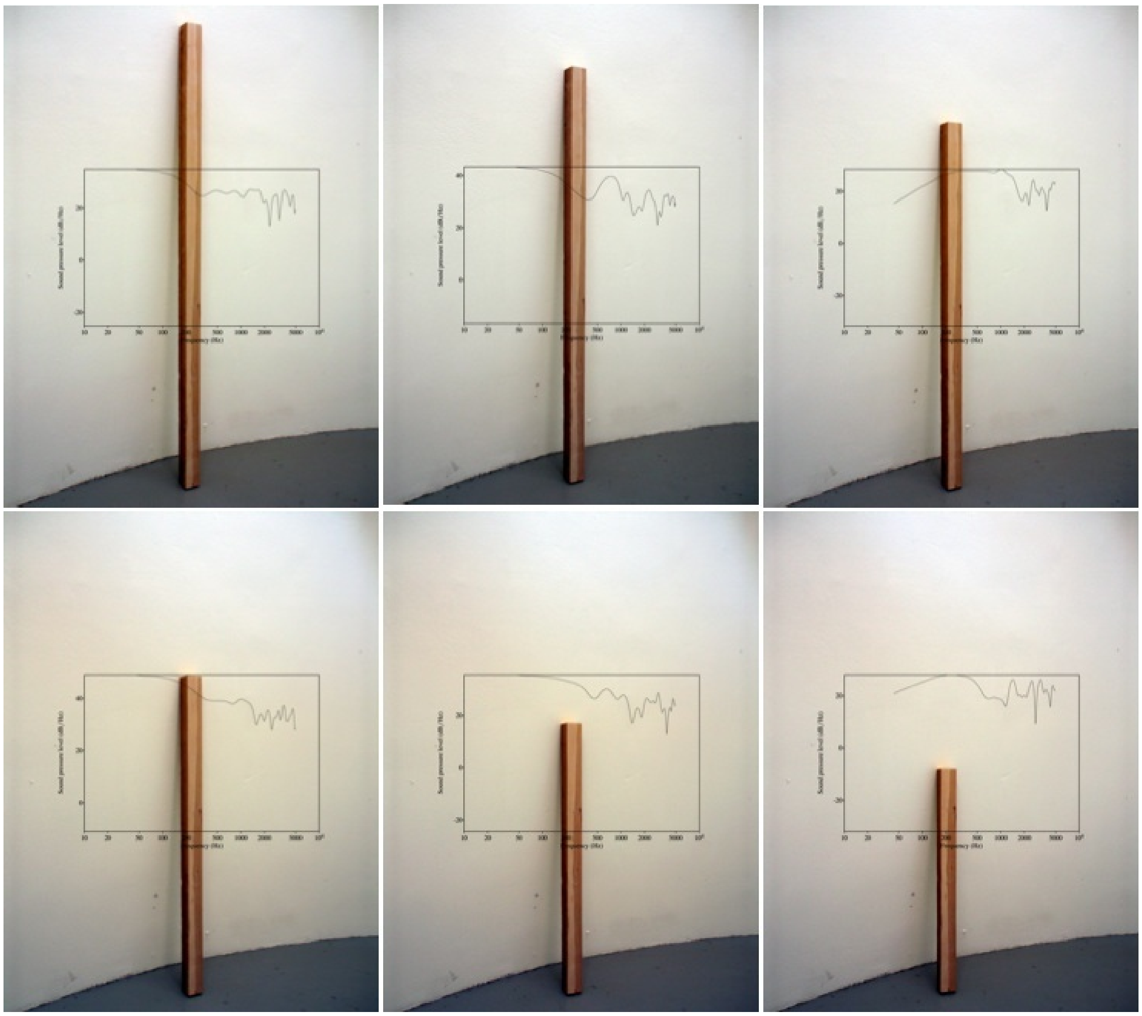


Figure 7.4: Cherry plank knocked with resultant power spectrums

looking into an analogue to speech recognition for the daxophone. The dax listening agent will be able to recognise specific gestures from the instruments and produce an output response of video clips of the plank knocked at its various lengths. We look forward to improvisations between the daxophones and the audio/visual traces of their source wood.

7.8 Future research: Visualisation of sound

Computer technology presently employed in psycho-acoustic science is inseparable from the questions that researchers in the field are inclined to ask and, in turn, the knowledge that the science generates. The recent proliferation of studies in the history of technology testifies to the interest in documenting this integration of technology and the knowledge it affords. In Psychology, this parallel is exemplified in Sigmund Freud's use of hydraulic metaphors to underpin a psychology of suppression and release. Accordingly, contemporary cognitive neuroscientists model mental processes on the dominant technology of our time: the computer.

Input/output models of cognition, and corresponding descriptions of "neural circuitry" are discursive tools that reflect the ubiquity of the computer, not only as a means for compiling and analysing behavioural data, but also as a model of cognition and a source of heuristic image. The history of computing, and computer image might constitute an archeology of thought patterns that have influenced the kind of experiments that get carried out, as well as the way behavioural data have been interpreted.

George Lakoff and Mark Johnson's foundational work on metaphors as constitutive of our experience of time and space will be an important text and reference for this study (Lakoff & Johnson, 1980). The work of Lera Boroditsky at Stanford will also be important, as her experiments suggest that conceptual metaphors not only condition our modes of thought, but also influence our perceptual systems (Boroditsky, 2001). Thus, this proposed inquiry will elaborate on some of the questions that emerge directly from this foundational masters work.

References

ANSI. (1994). *American national standard acoustic terminology.* New York.

Attneave, F., & Olson, R. (1971). Pitch as a medium: a new approach to psychophysical scaling. *American Journal of Psychology*, *84*, 147-166.

Auerbach, S. (1906). Zur lokalisation des musicalischen talentes im gehren unad am schadel. *Archives of anatomy and physiology*, 197-230.

Bharucha, J. J., Curtis, M., & Paroo, K. (2006). Varieties of musical experience. *Cognition*, *100*, 131–172.

Blesser, B., & Salter, L. (2007). *Spaces speak, are you listening*. Cambridge, Massachusetts: MIT press.

Boer, E. de. (1976). On the "residue" and auditory pitch perception. In *Handbook of sensory physiology* (Vol. V-3, p. 479-583). Berlin: Springer.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and english speakers' conceptions of time. *Cognitive Psychology*, *43*, 1-22.

Bregman, A. S. (1994). *Auditory scene analyses*. Cambridge, Massachusetts: MIT press.

Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, *22*, 349–368.

Burns, E., & Ward, W. (1978). Categorical perception - phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *Journal of the Accoustical Society of America*, *63*, 456-468.

Burns E, M., & Ward, D., W. (1978). Categorical perception - phenomenon or epiphenomenon: Evidence from the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, *63*, 456-468.

Butchel, H., & Butter, C. (1988). Spatial attentional shifts:implications for the role of polysensory mechanisms. *Neuropsychologia*, *26*, 499-509.

Cheveigné, A. de. (2004). Pitch perception models. In O. A. Plack Chris (Ed.), *Pitch: Neural coding and perception.* New York: Springer Verlag.

Clark, A. (1997). *Being there.* Cambridge, Massachusetts: MIT press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension.* Oxford, UK: Oxford University Press.

Cook, N. (1990). *Music, imagination, and culture*. Great Clarion Street, Oxford, UK: Oxford University Press.

Crowder, R. (1989). Imagery for musical timbre. *Journal of Experimental Psychology and Human Perception and Performance*, *3*, 477-478.

Dalla Bella, S., Giguère, J., & Peretz, I. (2009). Singing in congential amusia. *Journal of the Accoustical Society of America*, *126*(1), 414–424.

Delogu, F., Lampis, G., & Belardinelli, M. (2006). Music to language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers. *Cognitive Processing*(7), 203-207.

Denbigh, P., & Zhao, J. (1992). Pitch extraction and separation of overlapping speech. *Speech Communication*, *11*, 119-126.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, *3*, 357-370.

Du Verney, J. (1683). *Traité de l'organe de l'ouie, contenant la structure, les usages et les maladies de toutes les parties de l'oreille*. Paris: Paris.

Fourier, J. (1820). *Traite analytique de la chaleur*. Paris: Didot.

Galantucci, B., Fowler, C., & Turvey, M. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, *3*, 361-377.

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

Gibbs, R. W. (2005). *Embodiment and cognitive science*. New York, New York: Cambridge University Press.

Gockel, H., Moore, B., & Carlyon, R. (2001). Influence of the rate of change of frequency on the overall pitch of frequency-modulated tones. *Journal of Acoustic Society of America*, *109*, 701-712.

Goldstein, J. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Accoustical Society of America*, *54*, 317-317.

Goudbeek, M., Swingley, D., & Roel, S. (2009). Supervised and unsupervised learning of multidimensional acoustic categories. *Journal of Experimental Psychology: Human Perception and Performance*(6), 1913-1933.

Granzow, J. E., & Vokey, J. (2009). Musical expertise: Judging the pitch direction of implied fundamental frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, (accepted with forthcoming revisions).

Hanslick, E. (1957). *The beautiful in music*. New York: Liberal Arts Press.

Helmholtz, H. (1836/1954). *On the sensations of tone as a physiological basis for the theory of music* (A. Ellis, Ed.). New York: Dover.

Heylen, E., Moelants, D., & Leman, M. (2006). *Singing along with music to explore tonality.* Bologna: Alma Mater Studiorum, University of Bologna: 9th International conference on Music Percpetion and Cognition.

Hurley, S. (1998). *Consciousness in action*. Cambridge, Massachusetts: Harvard University Press.

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

Kaas, J., & Hackett, T. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology*, *9*, 164-170.

Kawahara, J.-I. (2007). Auditory-visual contextual cuing effect. *Percept Psychophys*, *69*(8), 1399–1408.

Krumhansl, C., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology and Human Perception and Performance*, *18*, 739-51.

Krumhansl, C., & Shepard, R. N. (1979). Quantification of the heirarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology and Human Perception and Performance*, *5*, 579-594.

Lakoff, G. (1989). The invariance hypotheis: Do metaphors preserve cognitive topology.

Duisburg.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lee, K. (2004). *Pitch perception: Place theory, temporal theory, and beyond* (Special Report). Stanford.

Leipp, E. (1977). *La machine à écouter: Essai de psycho-acoustique*. Paris: Masson.

Leman, M. (2008). *Embodied cognition and mediation technology*. Cambridge, Massachusetts: The MIT Press.

Levitin, D. (2002). Experimental design in psychological research. In D. Levitin (Ed.), *Foundations of cognitive psychology.* Cambridge, MA: MIT press.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 117-123.

Licklider, J. (1956). Auditory frequency analyses. In C. Cherry (Ed.), *Information theory* (p. 253-268). New York: Academic Press.

Liegeois-Chauvel, C., Peretz, I., Babai, M., Laguiton, V., & Chauvel, P. (1998). contribution of different cortical areas in the temporal lobes to music processing. *Brain*(121), 1853-1867.

Locke, J. (1690/1975). *An essay concerning human understanding*. Oxford: Clarendon Press.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action perception mismatch in tone deafness. *Current Biology*, *18*.

Macran, H. (1902). *The harmonics of aristoxenus*. Oxford: The Clarendon Press.

McAdamas, B., S. Giordano. (2009). The perception of timbre. In I. T. M. Hallam S. Cross (Ed.), *Oxford handbook of music psychology* (chap. 7). Great Clarendon Street, Oxford: Oxford University Press.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (2004). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Behavioral Science*, *58*, 177-192.

Mersenne, M. (1636). *Harmonie universelle*. Paris, Cramoisy: Editions du CNRS.

Meyer, J. (1978). The dependence of pitch on harmonic sound spectra. *Psychology of Music*, *6*.

Moore, B. (2003). *An introduction to the psychology of hearing* (5th ed.). London: Academic Press.

Neuhoff, J. (2004). *Ecological psychoacoustics: Introduction and history*. New York: New York : Academic Press.

Noë, A. (2004). *Action in perception*. Cambridge, Massachusetts: The MIT Press.

Noë, A. (2009). *Out of our heads: Why you are are not your brain, and other lessons from the biology of consciousness*. New York: Hill and Wang.

Ohm, G. S. (1843). Über die definition des tones, nebst daran geknüpfter theorie der sirene und ähnlicher tonbildener vorrichtungen. *Annalen der physik. Chem.*, *59*, 513–565.

Patel, P. (2008). *Music, language and the brain*. New York: Oxford university Press.

Patterson, R. (1990). The tone height of multiharmonic sounds. *Music-Perception*, *8*, 203-214.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain*, *132*, 1277–1286.

Pitt, M. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 976-986.

Plack, C., & Oxenham, A. (2005). The psychophysics of pitch. In R. Fay & A. Popper (Eds.), (Vol. 24, chap. 1). Springer New York.

Preisler, A. (1993). The influence of spectral composition of complex tones and of musical experience on the perceptability of virtual pitch. *Perception and Psychophysics*, *54*(5), 589-603.

Pressnitzer, D., Patterson, R., & Krumholz, K. (2001). Distortion products and the pitch of harmonic complex tones. In B. D.J., H. A.J.M., A. Kohlrausch, V. Prijs, & R. Schoonhoven (Eds.), *Physiological and psychophysical bases of auditory function* (p. 97-104). Maastricht: Shaker.

Rakowski, A. (1977). Memory for absolute and relative pitch. *Symp. Psychoacoustique Musicale, Paris*.

Rakowski, A. (2004). The music practitioner: Research for the music performer, teacher and listener. In J. Davidson (Ed.), (chap. 6). Ashgate: Aldershot.

Repp, B., & Knoblich, G. (2007). Action can affect auditory perception. *Psychological Science*, *18*, 6-7.

Rosner, B. (1988). Explorations in music, the arts and ideas. In E. . R. S. E. Narmour (Ed.), (chap. Music Perception, Music and Psychology). New York: Pendagron Press.

Ruggero, M. (1992). Responses to osound of the basilar membrane of the mammalian cochlea. *Current Opinion in Neurobiology*, *2*, 449 - 456.

Rusconi, E., Kwan, B., Giordano, B., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: The smarc effect. *Cognition*, *99*, 113-129.

Sano, H., & Jenkins, K. (1991). A neural network model for pitch perception. In P. Todd & G. Loy (Eds.), *Music and connectionism* (chap. 2). The MIT Press.

Sauveur, J. (1701). Systèm général des intervalles du son. In *Mémoires de l'academie royale des sciences.* Paris.

Schapin, S., & Schaffer, S. (1989). *Leviathan and the air-pump: Hobbes, boyle, and the experimental life*. Princeton, NJ: Princeton University Press.

Schlaug, G., Andrea, N., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Ann. N.Y. Acad. Sci*(1060), 219–230.

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

Schouten, F. (1938). The perception of subjective tones. *Proceedings of Koninkliijke*, *41*.

Schouten, J., Ritsma, R., & Cardozo, B. (1962). Pitch of the residue. *Journal of Acoustic Society of America*, *34*(1418 - 1424).

Schuppert, M., Munte, T., Wieringa, B., & Altenmüller, E. (2000). Receptive amusia: evidence for cross hemispheric neural networks underlying music processing strategies. *Brain*(123), 546-559.

Scruton, R. (1979). *The aesthetics of architecture*. Methuen, London: Princeton University Press.

Seebeck, A. (1841). Beobachtungen uber einige bedingungen der entste- hung von tonen [observations over some conditions of the emergence of tones]. *Annals of Physics and Chemistry*, *53*, 417– 436.

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., et al. (2007). Tone sequences with conflicting fundamnetal pitch and timbre changes are heard differently by musicians and nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 743-751.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music.* London: Academic Press.

Slevc, L., & Miyake, A. (2006). Individual differences in second language proficiency: Does musical ability matter. *Psychological Science*(17), 675-681.

Sloboda, J. (1992). *Toward a general theory of expertise* (K. A. Ericsson & J. Smith, Eds.). London, UK: Cambridge University Press.

Stainsby, T., & Cross, I. (2009). The oxford handbook of music psychology. In S. Hallam, I. Cross, & M. Thaut (Eds.), (chap. 5). Oxford University Press.

Stalinski, S., Schellenberg, E., & Trehub, S. (2008). Developmental changes in the perception of pitch contour: distinguishing up from down. *Journal of Acoustic Society of America*, *124*, 1759-1763.

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

Stewart, L. (2008). Do musicians have different brains. *Medicine, Music and the MInd*, *8*, 304-308.

Terhardt, E. (1974). Pitch, consonance and harmony. *Journal of the Accoustical Society of America*, *55*, 1061-1069.

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

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

Turner, M. (1990). Aspects of the invariance hypothesis. *Cognitive Linguistics*, *1*, 254.

Turner, R. (1977). The ohm-seebeck dispute, hermann von helmholtz, and the origins of physiological acoustics. *The British Journal for the History of Science*, *10*, 1-24.

Ueda, K., & Ohgushi, K. (1987). Perceptual components of pitch: Spatial representation using a multidimensional seeing technique. *Journal of the Acoustical Society of America*, *82*, 1193-1200.

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

Watts, C., Moore, R., & McCaghren, K. (2005). The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *Journal of Voice*, *19*, 534-543.

Watts, C., Murphy, J., & Barnes-Burroughs, K. (2003). Pitch matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with nontalented singing voices in conditions of varying feedback. *Journal of Voice*, *2*, 185-186.

Wightman, F. (1973). The pattern-transformation model of pitch. *Journal of the Accoustical Society of America*, *54*(2), 407-416.

Yates, G., Winter, I., & Robertson, D. (1990). Basilar membrane non-linearity determines auditory nerve rate-intensity functions and cochlear dynamic range. *Hearing Research*, *45*, 203-220.

Zarate, R. J., J. M.and Zatorre. (2008). Experience-dependent neural sub- strates involved in vocal pitch regulation during singing. *Neuroimage*, *40*, 1871-1887.

Zbikowski, L. M. (2002). *Conceptualizing music*. New York, New York: Oxford University Press.

Zemp, H. (1981). Melanesian solo polyphonic panpipe music. *Ethnomusicology*, *25*, 383-418.