1-1-2009

Feel the music : crossmodal integration in music perception

Michael Maksimowski
Ryerson University

FEEL THE MUSIC: CROSSMODAL INTEGRATION IN MUSIC PERCEPTION

by

Michael Maksimowski

Hon. B.Sc., University of Toronto, June, 2007

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Masters of Arts

in the Program of

Psychology

Toronto, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis or dissertation.

I authorize Ryerson University to lend this thesis or dissertation to other institutions or individuals for the purpose of scholarly research.

—

I further authorize Ryerson University to reproduce this thesis or dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Abstract

Michael Maksimowski

Feel the Music: Crossmodal Integration in Music Perception

M.A. Psychology

Ryerson University

Toronto

2009

In addition to auditory information, music perception often involves visual and vibrotactile information, making it an ideal domain through which to study cross-modal integration. Recent research has demonstrated a strong influence of visual information on auditory judgments concerning music. However, we have very little empirical information regarding integration of vibrotactile information in music. In Experiment 1, participants made judgments of interval size for unimodal presentations of melodic intervals in auditory, visual, and vibrotactile conditions. In Experiment 2, participants made judgments of interval size for cross-modal presentations of intervals comprised of stimuli presented in the three unimodal conditions of Experiment 1. In Experiment 3, participants were trained with vibrotactile stimuli to assess if learning benefits audio-vibrotactile integration in music perception. The results are discussed in light of differences in the extent of visual and vibrotactile influence on auditory judgments and the role of learning in cross-modal integration in music.

Acknowledgements

The author would like to acknowledge the technical assistance of Gabe Nespoli.

Dedication

To my parents who have always given me the love and support to accomplish everything I have accomplished.

Table of Contents

List of Tables

List of Figures

List of Appendices

Feel the music: Crossmodal integration in music perception

Music is an art and a form of communication (Mithen, 2005). Although the auditory dimension is typically the most salient, visual and tactile dimensions of music are important but often-neglected dimensions of performance. In many respects, these dimensions serve to enhance not only the entertainment value from perceiving music, but also music information. Visual and tactile information such as facial gestures, body movements, and vibrations convey timing, perceptual, and affective information. While our understanding of visual influences on music perception is increasingly well known (Thompson, Graham, & Russo, 2005), the same cannot be said for the tactile modality.

If we consider music as a crossmodal stimulus and the perceptual benefits gained from audio-visual integration, it is of interest to investigate whether similar potential benefits can be obtained from using tactile information. The present study assessed whether interval estimates (a necessary skill for interpreting melody) can benefit from a vibrotactile medium. Furthermore, the present study also investigated whether prior learning necessitates such benefits, if any. The following hypotheses were tested: (1) that vibrotactile signals can improve accuracy for interval size estimates above chance, (2) that vibrotactile signals paired with auditory signals can enhance accuracy for interval size just as much as visual signals paired with auditory signals can, and (3) that training can further improve the benefits from vibrotactile signals towards estimates of interval size.

The Concept of Crossmodal Perception

Perception is fundamentally a multisensory experience; sight, hearing, touch, taste, and smell can rarely be construed as pure, independent modalities but rather as modalities

that reciprocally influence each other. Depending on the nature of the stimuli involved, perception can be enhanced or reduced when multiple modalities capture information attributable to a particular objective. Additionally, cognitive interpretations and behavioural responses towards a specific stimulus change based on the context in which the stimulus is perceived. For example, our interpretation of a song differs depending on whether we listen to it on a CD or watch and listen to it on television; the auditory component is similar between situations and it is therefore the visual component that changes the musical interpretation.

Crossmodal integration refers specifically to the effect by which two distinct stimuli are perceived as emanating from the same object or situation; This leads to a disambiguation of object identity, purpose, and/or meaning to the perceiver (Meredith & Stein, 1986). Only in recent years have psychologists and neuroscientists begun to understand how multisensory perceptions influence cognition and how sensory channels interact with one another, respectively (Calvert, Spence, & Stein, 2004). Indeed, the current debate is not whether crossmodal integration occurs, but at what processing stage modalities integrate percepts. The widely held theory among researchers today is that percepts are immediately removed from their generators (i.e., modalities) and integrated into high-level crossmodal representations (e.g., Molholm, Ritter, Murray, Javitt, Schroeder, & Foxe, 2002). Upon integration, the unimodal cues no longer become distinguishable; crossmodal neurons act to transmit sensory information into an integrated product that carries a whole new meaning to higher-level processing areas.

Theoretical Considerations

The advantages of crossmodal integration are both fascinating and perplexing to researchers. Crossmodal perception is fast, natural, automatic, effortless, and pre-attentive. This is not to say that unimodal perception is non-existent. Rather, it is advantageous to utilize multiple modalities in order to reduce perceptual ambiguity (MacLeod & Summerfield, 1990) improve stimulus detection (Lovelace, Stein, & Wallace, 2003; Stein, London, Wilkinson, & Price, 1996; Vroomen & de Gelder, 2000), and reduce reaction times (Driver & Spence, 1998; Hershenson, 1962; Spence & Driver, 1996). Although crossmodal integration can lead to errors of perception in unusual circumstances, the advantages previously noted vastly outweigh the disadvantages. It is important to note that crossmodal perception does not involve making cognitive compromises or summations between differing modalities; crossmodal signals change what the listener perceives altogether (Liberman, 1984).

One particular question of interest is whether crossmodal integration is an innate or learned process. Two general accounts of crossmodal integration can be derived from current theories of perception. In the first account, perceivers consult memory representations that include specific examples of modalities currently being utilized (Massaro, 1987, 1989). For example, listening/watching a speaker talk activates memory representations of fundamental auditory and optic units of spoken utterances. In this theory, prototypes are formed that the perceiver is able to utilize for making interpretations. In the cases where the modalities are inconsistent with one another, the perceiver selects and experiences the prototype most consistent with the collection of cues. Thus, visual and tactile modalities influence the sound percept based on the association of optical-acoustic

and tactile-acoustic cues in memory, respectively. These associations, in turn, are derived from both natural and symbolic cue associations sampled by perceptual modalities.

In the second account, no memory representation is necessary in order to perceive stimuli from our environment. The motor theory suggests that listeners perceive the significant gestures (or origins) of the object that produced the modal signal rather than the modal signal *per se*. Indeed, a closer interaction may exist between the listener's own percepts and the object's gestures (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). The intentionality and underlying movements necessary to generate the signal are of utmost importance. Motor theorists in speech research argue that greater correspondence exists between the listener's percept and the speaker's vocal-tract gestures than between the listener's percept and the acoustic signal. Perceivers thus devise hypotheses using both the signal and source rather than learned associations.

Other researchers (Gibson, 1966, 1979) agree with the motor theory in much the same manner, except that recovery of the percepts is considered to be even more immediate. Perception of the signals themselves is unnecessary and fallible. Within the direct-realist theory, perception is seen as the only means by which individuals can understand their environment. Because of this, a direct link exists between one's perceptual systems and the environmental media. Individuals thus perceive environmental sources rather than the signals themselves. For example, listeners perceive the causes of an acoustic and visual signal rather that the actual combined signal.

Although direct-realism suggests that we carry an immediate (direct) perception of objects, qualities, and events, it is false to suggest that this theory also claims we see the

world exactly how it is. Direct-realism does not deny perceptual illusions or misinterpretations of the world. However, when we use our perceptual systems, direct-realists claim that our object of perception is in the external world rather than the mind. In other words, perceptual benefits from crossmodal integration are not restricted to prior experiences.

The above-mentioned theories have gained variable levels of acceptance and support from researchers, with no theory particularly winning out. The motor theory is criticized for carrying no explanation as to how the speaker's vocal-tract gestures are translated into a percept for the listener. The direct-realist theory fails to fully explain mental processes such as hallucinations, dreams, and imagining. Such processes activate visual and auditory perceptions of images rather than external objects. Finally, the prototype model faces two main objections: (1) we have an insufficient memory capacity to store every percept and (2) the percepts produced by the environment may not even be similar to our stored prototypes, suggesting that perception cannot be explained by this theory alone. Although it remains to be determined whether one theory can best account for crossmodal perception, the following study is not a direct test of theories. Rather, these theories have been presented in order to provide one of the motivational factors for studying a relatively unexplored crossmodal pairing in music (auditory-vibrotactile).

As previously discussed, music can be perceived using auditory, visual, and tactile senses. In the following subsections, two crossmodal pairings will be described: auditory-visual and auditory-tactile. These pairings will then be discussed within a musical context. I argue that our experience with auditory-visual integration in music is much greater relative to our experience with auditory-tactile integration in music. If crossmodal integration in

music is not dependent on learning, then a relatively unfamiliar modal pairing (auditory-tactile) should still provide benefits superior to a unimodal presentation (audio). However, if crossmodal integration in music is dependent on learning, then training on an unfamiliar unimodal stimulus (tactile) and/or training on an unfamiliar modal pairing (audio-tactile) should be necessary before any perceptual benefits are received.

Audio-Visual Integration

Audio-visual integration is everywhere. Movies, concerts, and conversations are just some of the many examples in which we integrate auditory and visual information into one perception. Indeed, sounds are identified not only by our ears, but by our eyes as well. While the auditory system is adept at first specifying a general search area for a sound, the visual system is capable of pinpointing the source after the general location is given. Our visual system can also help disambiguate what we are hearing. For example, seeing and hearing a speaker can improve the listener's understanding of both speech content and intention (Büchel, Price, & Friston, 1998).

Two excellent examples of auditory-visual integration are the Ventriloquist Effect and McGurk Effect. The Ventriloquist Effect (Howard & Templeton, 1966) is an umbrella term for all forms of spatial auditory-visual integration. When the perceived location of a sound differs from the actual sound source, such as when watching television and movies, the visual source "captures" the sound so that we actually hear voices, music, and sound effects coming from the video source itself. The McGurk Effect is a compelling illustration of how visual speech is automatically integrated into what we hear (McGurk & MacDonald, 1976). In this phenomenon, a participant watches a video in which the phoneme produced is "ga". However, the sound is dubbed over with the phoneme "ba". Listeners end up

perceiving the phoneme "da". Thus, the visual phoneme fundamentally changes what the listener hears. This effect occurs in pre-linguistic infants (Rosenbaum, Schmuckler, & Johnson, 1997), with whole words (Dekle, Fowler, & Funnell, 1992), and even after extensive training (Massaro, 1987). These phenomena are just two of many examples that showcase the extent to which visual information influences audition.

Learned associations are most likely to be found between auditory and visual modalities due to their frequent pairing in everyday life. Individuals have ample opportunity to develop memory representations that include visual and auditory cues. However, crossmodal perception of an event can be seen as a natural association; if perception captures environmental sources, as proponents of the motor and direct-realist theories suggest, different media that co-occur within the same event serve to inform the perceiver in a joint fashion. Due to our extensive pre-existing exposure to auditory-visual stimuli, this crossmodal pairing alone is not suitable to address whether a learning component exists for crossmodal integration. It is necessary to also utilize a crossmodal pairing that is relatively "unlearned".

Audio-Tactile Integration

Tactile perception refers to a noticeable change in mechanical pressure or distortion. The majority of tactile sensors (mechanoreceptors) are located on the skin, while other small yet sensitive groups are found in the vestibular system (for balance and spatial orientation) and cochlea (auditory perception from air pressure waves). Due to a shared energy source, all sound consists of vibrations and all vibrations consist of sound, making a congruent pairing between auditory and tactile information lawfully related. For example, the beat of a drum produces changes in pressure in both the drum and through the air. Both

of these pressure changes originate from a common environmental event (i.e., the drum), making such an association lawful by nature. It remains debatable however whether auditory-tactile pairings are experienced to the same extent as auditory-visual pairings.

Fowler & Dekle (1991) argue that audio-tactile associations, such as hearing and feeling the productions of a speaker's vocal tract, are lawful in the sense that both modalities capture changes in pressure; the auditory system transduces changes in air pressure into sound and the tactile system transduces changes in material pressure into touch. The authors also suggest that audio-tactile associations are relatively unfamiliar to the average person because we are not typically exposed to the tactile properties of the vocal cord when understanding what someone is saying.

One striking example of auditory-tactile integration is the Tadoma Effect, in which tactile speech is automatically integrated into what we hear (Fowler & Dekle, 1991). This is a phenomenon similar to the McGurk Effect and distinct from the Tadoma Method, in which individuals who are deaf or hard of hearing place their thumb and forefingers on the lips and vocal cords of the speaker, respectively, in order to perceive tactile properties of speech (Alcorn, 1932). The Tadoma Effect occurs when auditory perceptions of "ba" are dubbed with tactile perceptions of "ga", creating the end-perception of hearing and feeling "da". This effect works even when the phonemes are switched between modal signals. As will be discussed shortly, Fowler & Dekle (1991) used the Tadoma Effect to address competing perceptual theories (described above) behind the McGurk Effect.

One hypothesis in crossmodal speech research is that auditory-tactile pairings are not necessarily learned to the same extent as auditory-visual pairings (Fowler & Dekle, 1991; Gick, Jóhannsdóttir, Gibraiel, & Mühlbauer, 2008). The degree to which we

experience auditory-tactile pairings, however, is not entirely unfamiliar. Music experiences at concerts, clubs, and movie theatres typically involve the sensation of bass and subwoofer vibrations presented along with the music. Bone vibrations contribute to sound and self-speech perception (e.g., Sohmer, Freeman, Geal-Dor, & Savion, 2000). Self-speech perception involves vibrotactile feedback via bone conduction (Shuster & Durrant, 2003). However, these experiences are restricted to low-end frequencies, which constitute a minor percentage of the frequencies typically experienced from music. Békésy (1949) found that there were great losses in intensity of one's own speech (via bone conduction) at high frequencies compared to low frequencies. These low-end frequencies (10 – 250 Hz) also constitute the majority of vibrations experienced in large settings where music is heard. Indeed, the majority of commercial subwoofers utilize a frequency range from 20-150 Hz, while industrial subwoofer systems use an even lower range (20-80 Hz). Auditory frequencies from music can range from 20 to 4400 Hz. In some situations, materials and objects within the musical environment (such as a chair or table) can resonate from the auditory stimuli being produced. However, the low-end frequency bass masks and distorts higher frequency vibrations. Thus, humans are not very experienced with a complete mapping between auditory and tactile music frequencies relative to our mapping between auditory and visual music frequencies.

Tactile Enhancement of Speech: Evidence of Auditory-Tactile Integration

Speech researchers are particularly interested in understanding crossmodal integration. Fowler and Dekle (1991) sought to determine whether a crossmodal phenomenon (the McGurk Effect) was due to learned associations or whether more direct associations between percept and sources could account for the effect. In order to

accomplish this, conditions were fashioned in such a way so that judgments within each condition could be attributed to either learned or relatively unlearned associations. Participants listened to a continuum of 10 different syllables, ranging from high frequency /ba/ to low frequency /ga/, and were asked to report whether they heard "ba" or "ga". The dependent variable was the extent to which participants reported hearing and feeling "ba" or "ga" judgments across the continuum.

In one condition, participants were presented with typed visualizations of /ba/ and /ga/ on a computer. Hearing "ba" and seeing an orthographic transcription of /ba/ does not represent a natural coexistence in the environment, but rather a societal convention. In a separate condition, participants felt the vocal tract of a speaker who either uttered "ba" or "ga" on each trial. This procedure mimics the Tadoma Method (described above).

Acoustic and tactile stimuli can co-arise from the same environmental event, thus making this association a lawful pairing. Additionally, no subject reported having any training or extensive experience in which they felt someone's face at the same time that person was talking. Although this does not single out infantile experiences, experience with auditory-tactile pairings of such a manner would still be considerably less than auditory-visual pairings of speech.

The authors predicted that if the memory representation theory of perception is correct, the orthographic condition should show a McGurk-like effect, while the Tadoma condition should not. If the motor or direct-realist theories of perception are correct, however, the Tadoma condition should show a McGurk-like effect, while the orthographic condition should not. A McGurk-like effect would be confirmed if feeling and/or hearing "ba" were significantly higher than feeling and/or hearing "ga" at the /ga/ end of the

continuum. Alternatively, neither or both conditions could show McGurk-like effects, implying that crossmodal integration is both innate and acquired through memory representations.

Results showed that the orthographic condition had no crossmodal effect, while the Tadoma condition did. Specifically, significantly higher reports of "ba" versus "ga" were made at the /ga/ end of the auditory continuum when /ba/ was also felt. The Tadoma Effect was also noticed immediately in the first block of trials. Interestingly, not only did felt syllables significantly affect judgments of the syllable heard, but the acoustic syllable significantly affected judgments of the syllable felt. Thus, considerable integration of information from the two modalities must have taken place. Similar results were found even when the experimenters equalized attentional demands between conditions.

In a separate study, Gick, Jóhannsdóttir, Gilbraiel, and Mülbauer (2008) created two bimodal conditions to determine the relative contributions of tactile information to both visual and auditory speech signals. A trained experimenter produced all disyllables (a word comprising of two syllables) perceived during the experiment. Within the auditory-tactile condition, participants listened through headphones with white noise and felt the disyllables from the experimenter using the Tadoma Method. Participants were also instructed to close their eyes to eliminate any visual information from being perceived. Within the visual-tactile condition, auditory information was eliminated through white noise played through headphones; Participants were instructed to use the Tadoma Method and watch the speaker as she mouthed the disyllables. Auditory-alone and visual-alone conditions served as controls. The dependent variable in all conditions was whether the participant was able to correctly repeat the disyllable they just perceived.

Accuracy in both the visual-tactile and auditory-tactile conditions was significantly higher than their respective controls. Specifically, tactile information improved accuracy by about 10% in both bimodal conditions. However, tactile signals provided similar amounts of information to both auditory and visual speech signals. Interestingly, a negative correlation was found in which participants who gained greater benefits from tactile information for understanding visual speech gained far less benefit from the same information for understanding auditory speech and vice-versa. These two studies are promising illustrations of how relatively unfamiliar crossmodal signals can benefit speech recognition. These advantages may extend to other stimuli such as music.

Music as a Crossmodal Experience

Music acts as a useful stimulus for studying crossmodal integration in many ways. First, music shares a number of similarities with language, a form of communication that has already been extensively studied from a crossmodal perspective (Büchel, Price, & Friston, 1998; Granström, House, & Karlsson, 2002; McGurk & MacDonald, 1976). Language and music can be thought of as forms of communication, with the former primarily delivering information and the latter conveying intention and emotion (Mithen, 2005; Patel, 2007). Similar to language, music carries syntactical structure with the organization of notes, rhythms, phrases, chords, and keys (Lerdahl & Jackendoff, 1983). Second, visual and tactile percepts inherent in music productions integrate with auditory percepts to produce an entirely different musical experience. Facial gestures, body movements, onset/offset cues, and vibrations convey timing, perceptual, and affect information about the music (Thompson, Graham, & Russo, 2005).

Third, music perception can be seen as a highly ambiguous perceptual experience that consequently lends itself to significantly large gains from crossmodal integration. When experiencing music, the listener must make sense of melodic and temporal organization, emotional intent, timbre, facial/body language, intensity, and so on. Thus, music perception has the potential to benefit greatly from crossmodal integration. According to the inverse effectiveness rule, the greater the ambiguity within the unimodal signals, the more effective they are in combination in terms of identifying and/or locating an object (Stein, Meredith, & Wallace, 1994). In other words, crossmodal integration is greatest when a stimulus is poorly identified using unimodal signals. Indeed, multisensory neural responses are greatest when the neural responses to each unimodal stimulus are smallest (Stein & Meredith, 1993). Alais and Burr (2004) presented unimodal and bimodal visual and auditory stimuli to participants, instructing them to localize the stimulus or stimuli presented to them. When the visual signal was clearly displayed on a bimodal trial, vision dominated judgments of location irrespective of where the sound signal originated (a Ventriloquist-like effect). When the visual stimulus was distorted on a bimodal trial, hearing dominated judgments of location irrespective of where the visual signal originated. Most importantly, however, when visual and auditory signals were equally distorted, neither sense dominated. This resulted in significantly lower errors in location judgment compared to either unimodal signal alone.

Music research has generally focused on auditory aspects such as pitch, timbre, and dynamics without necessarily determining the relative contributions that vision and touch can have on what the listener perceives. Indeed, these visual and tactile features have been separated from the audio with the introduction of recording technologies such as the radio and portable music players. From this, our conception of music has been altered so that

visual and tactile information is often downplayed or ignored in both everyday life and in research (Thompson & Russo, 2005). This is not to say that these features are unwanted. In fact, they are likely a main reason why live music concerts and performances are still highly desired (not discounting social factors as well). Researchers have recently begun to consider crossmodal contributions to music performance and experience. Dissanayake describes the evolution of music as a 'multimedially presented and crossmodally processed activity of temporally and spatially patterned vocal, bodily, and facial movements' (Dissanayake, 2001: 389), while Cook (1998) theorizes about the crossmodal experiences inherent in ballet, opera, and music video clips.

Visual Music. As previously discussed, the influences of visual features on music perception have already been studied. Visual perception of music performances greatly influences how the music is perceived and understood. For example, facial movements of musicians and singers can influence our perceptions of interval size. In the absence of sound, Thompson and Russo (2007) presented videos of ascending melodic intervals being sung by trained female vocalists. Mean ratings of perceived interval size were significantly correlated with the extent to which the singers displaced their head, eyebrows, and mouth when transferring from the low note to high note. The actual interval sung also showed a similar strong positive correlation with the displacement of these visual features.

Visual signals from musicians have been shown to influence ratings of affect. Gestures, facial expressions, and body language all convey visual information that influences one's emotional judgment of auditory stimuli (Thompson, Russo, & Quinto, 2006). Specifically, visual features direct the listener's attention to emotional content (Davidson & Correira, 2002) and specifically in the case of music, components such as

rhythm, harmony, and melody (Thompson, Graham, & Russo, 2005). Juslin (2001) suggests that emotional facial expressions are inherent in musical performances. Thompson and Russo (2004) presented audio clips of sung major and minor intervals that are typically perceived as happy and sad, respectively. In addition to the audio, a visual component was added that was either congruent or incongruent with the audio clip. Judgments of emotional meaning were significantly affected by audio and visual information, with no interaction between the two signals. The judgments between auditory and visual information were found to be additive. Even though participants were instructed to make their judgments from the music alone, visual features had a greater influence on emotional ratings than audio features.

Visual gestural movements of musicians also influence music perception. Vines, Krumhansl, Wanderley, and Levitin (2006) asked trained musicians to hear, watch, or hear and watch clarinetists perform a musical piece. Results revealed that visual information had both an augmentation and reduction effect on the participants' experience of tension at different points in the musical piece. The researchers argued that visual features such as arm/head movements and facial expressions indicated structural features such as phrasing and interpretive features such as emotional content. In a separate study, Saldaña and Rosenbaum (1993) showed that individuals "hear" a note begin more abruptly when they see a string player pluck their instrument versus bowing it. Thus, a musician's motor movements (face or body) can convey information that is both relevant and influential towards how the music is ultimately understood.

Tactile Music. The benefits of tactile components towards music perception are not well known. This is surprising considering that tactile signals naturally co-occur whenever

music is played. Indeed, sound and vibrations capture the same energy source, albeit from different mediums (air pressure for sound and material pressure from vibration). A number of studies by Marcelo Wanderley and colleagues have demonstrated the important role of vibration in music production (e.g., Bimbaum & Wanderley, 2007; Marshall & Wanderley, 2006). For any musician, sound and touch are tightly entwined. In fact, tactile feedback is seen as essential towards playing any instrument; feeling the vibratory, thermal, and textural properties of the instrument (with the hands and on the mouth for instruments with mouth pieces) gives the performer a better understanding of what he or she is playing and the ability to make playing adjustments due to the feel of the instrument rather than sound alone (Howard, Rimell, Hunt, Kirk, & Tyrrell, 2002). Research on how vibrations can influence music perception is growing. For example, music presented as vibrotactile stimulation can invoke emotions consistent with those experienced with sound (Karam, Branje, Russo, Price, & Fels, 2008; Karam, Nespoli, Russo, & Fels, 2009).

The Present Study

To date little research has been conducted on the extent to which listeners can gain useful musical information from vibrotactile signals. This may be because we are unfamiliar with how vibrations work in music and therefore do not consider them an aid for musical judgments. The goal of this study was to investigate the following questions: To what extent can vibrations augment the perception of musical stimuli? Is crossmodal integration in music dependent on learning (i.e., prior exposure and familiarity with the congruent presentation of two or more modal signals)? As previously discussed, significant audio-visual associations have already been found for both language and music. These associations, however, can be learned and/or innate. The question remains whether similar associations can also be found between auditory and vibrotactile signals within a musical

context. Specifically, can vibrotactile signals supplement relevant information about the music?

Three separate experiments were designed to investigate these questions. The first experiment looked at the relative contributions from single modalities (audio, video, vibrotactile) towards understanding musical information. I believe this to be an important consideration because if we are to assume that visual and vibrotactile signals help us to better understand the auditory music stimulus, then there should be some capacity for relative judgment benefits from these signals alone. The second experiment utilized audio-visual and audio-vibrotactile stimuli to assess whether congruent crossmodal presentations could enhance music judgment over audio-alone. More specifically, I was interested in assessing whether audio-vibrotactile stimuli could produce not only enhanced judgments over audio-alone stimuli, but similar levels of accuracy as when audio-visual stimuli were used. The third experiment served as a follow-up to the second experiment and looked at whether minimal training and modal mapping could improve musical judgments.

Measures

Interval Size Judgments. The perception of interval size is closely related to relative pitch, our sensitivity to relations between pitches. Relative pitch is a universal human ability that is necessary for understanding speech prosody and musical melodies. Judgment of interval size is a useful task for objectively measuring a participant's ability to perceive relative pitch. Russo and Thompson (2005) showed that although musicians and non-musicians experience different degrees of differentiation for intervals within an octave, mean estimates of interval size for both groups were highly correlated with log frequency

distance. This finding suggests that individuals can scale interval size regardless of musical background.

Musical expertise or training has been suggested to influence estimates of interval size. Siegel and Siegel (1977a) showed that musicians are resistant to the influence of context when judging interval size, unlike untrained participants. In other words, musicians appear to acquire categories for pitch that have a functional similarity to phonemic categories for speech. Musicians are also adept at categorical perception of musical notes (Siegel & Siegel, 1977b). Thus, musical background is an important consideration when assessing an individual's capability of estimating interval size.

Interval sizes were measured using semitones, with the smallest and largest interval sizes being 0 and 12 semitones, respectively. Participants would judge the interval size on each trial and were given practice trials before any testing was conducted. Estimates of interval size were used to calculate the dependent variables of accuracy and precision, as discussed below.

Accuracy. Scores were assessed based on how close each participant was able to guess the correct semitone range (i.e., the absolute difference between the actual semitone range and the perceived pitch range in semitones). For example, if a participant estimated "10 semitones" or "6 semitones" on a pitch interval that actually ranges 8 semitones, a score of 2 would be given for either response. Score was inversely related to accuracy, with lower scores indicating better accuracy and higher scores indicating poorer accuracy.

Precision. Another factor of interest was the level of reliability in participants' scores for each unique trial. Precision is independent of accuracy and is a reflection of variability of judgments. Thus, higher precision indicates lower variability and suggests that

a participant was more consistent in his or her responses. Precision was calculated for each interval using the standard deviation of participant responses for a particular interval within a condition. Once again, score was inversely related to precision, with lower scores indicating better precision and higher scores indicating poorer precision.

Questionnaire. All participants filled out a questionnaire upon completion of testing. The self-made questionnaire included personal information such as the participant's age, sex, handedness, first language, and hearing ability. Additional questions concerned the participant's musical expertise (formal and informal) and musical preferences. Musical expertise served as a covariate based on the total number of years the participant had engaged in formal vocal, instrument, and/or music theory training, regardless of the participant's age at the time of training. See Appendix E for the actual questionnaire.

Experiment 1: Unimodal Presentations

The following experiment investigated the extent to which interval size could be judged from three separate unimodal signals: audio-alone, visual-alone, and vibrotactile-alone. While humans are adept at making judgments of interval size using auditory information alone, it is unknown to what extent judgments of interval size can be made using just visual or vibrotactile modalities.

I predicted that accuracy scores for all 3 conditions would be greater than chance (see Appendix C for how chance estimates were calculated). Most individuals have extensive experience listening to music; the same cannot be said of only watching or feeling music. Therefore, the auditory condition was also expected to yield the highest accuracy levels compared to the visual and vibrotactile conditions. Given that participants have more experience and familiarity with visual music than vibrotactile music, I also predicted accuracy in the visual condition to be better than the vibrotactile condition.

Methods

Participants. A total of 33 Ryerson undergraduate students (average age = 23.1 years, 4 male, 2 left-handed) with an average 4.1 years of musical training (*SD* = 4.8) participated for course credit. Data was analyzed only from participants who had normal hearing (hearing level was reported on the questionnaire).

Test Stimuli. Two female amateur singers were paid to sing ascending intervals that ranged from 0 to 12 semitones (Table 1). A note range of 220-440 Hz was used which contained the optimal frequency range of vibrotactile sensitivity, 250-300 Hz (Gescheider, Bolanowski, Pope, & Verillo, 2002; Watson, 1979), and the peak frequency output (250 Hz) of the vibrotactile source (discussed below). Twenty-six trials were selected as stimuli, 22 of which were used for testing (all possible diatonic intervals spanning 1-11 semitones from

both singers) and 4 for practice (diatonic intervals of 0 and 12 semitones from both singers). Audio and video were recorded from each singer's performance and all notes were sung with the syllable "la". From these media files, three different modal signals were created: auditory, visual, and vibrotactile. Each media file contained two sung notes: The first and second note were sung for approximately 1.5 seconds each, with a 0.5 second pause in-between the 2 notes. Each trial lasted for an average of 4.43 seconds (*SD* = 0.19). A range of an octave for the musical stimuli was chosen for musical relevance; melodic intervals greater than an octave in separation are extremely rare in music.

Table 1.
List of semitone ranges used and their respective notes/frequencies (displayed in Hz).

Semitone Range	First Note		Second Note		Frequency Difference
	Note	Frequency	Note	Frequency	
0	Eb4	311.5	Eb4	311.5	0
1	Eb4	311.5	E4	329.65	18.15
2	D4	293.65	E4	329.65	36
3	D4	293.65	F4	349.25	55.6
4	C#4	277.2	F4	349.25	72.05
5	C#4	277.2	F#4	370	92.8
6	C4	261.65	F#4	370	108.35
7	C4	261.65	G4	392	130.35
8	B3	245.95	G4	392	146.05
9	B3	246.95	G#4	415.3	168.35
10	Bb3	233.1	G#4	415.3	182.2
11	Bb3	233.1	A4	440	206.9
12	A3	220	A4	440	220

Two experts helped calibrate the auditory and vibrational intensity for each trial. The zero semitone trial from the first singer served as a comparison to all other trials. See Appendix B for an assessment of non-calibrated stimuli.

Design. Each experiment was conducted in a double-walled sound attenuation chamber (IAC). Participants sat in a chair facing a table so that they looked away from the window. The doors of the chamber were closed before testing began.

Figures 1a, 1b, and 1c. Experimental set-up, video frame from a trial, and Tactaid skin transducers.

Auditory stimuli were played over Sennheiser HD535 headphones at a comfortable listening volume. Visual stimuli were displayed on a 13-inch laptop (MacBook, Apple) placed approximately 50cm away from and directly in front of the participant. Visual information for any trial consisted of the singer's face, neck and shoulders. All videos appeared in the centre of the screen against a white background. Vibrotactile stimuli were presented through 2 skin transducers (VBW32 Tactor, Audiological Engineering Corporation) with a peak frequency of 250Hz, a transient response of 5ms, and a nominal output of 100-800Hz. The Tactaid is a lightweight voice coil device primarily manufactured as a speech aid. These transducers were chosen because of their small size and ability to be easily placed in the hands of the participant. The hands were chosen for perceiving vibrations because they contain the greatest numbers of Pacinian Corpuscles, a type of mechanoreceptor that is selectively responsive to vibration. Participants were instructed to place a transducer in each hand, wrap their fingers around it, and rest their hands either on the table or on their lap during testing (except when they had to make a response). Both transducers produced the same signal and were attached to an amplifier (HP4, PreSonus).

Participants were given unimodal presentations of the musical intervals. There were 3 within-subjects conditions. In the auditory-alone (A) condition, participants only heard the semitone intervals sung. In the visual-alone condition (V), participants only viewed videos

in which the pitch intervals were sung. In the vibrotactile-alone (T) condition, participants only felt the pitch intervals sung. Each condition consisted of 66 trials (22 unique trials repeated 3 times for 3 different conditions for a total of 198 trials) with an equal number of trials allocated to each singer and interval. Two different between-subjects condition orders were used: A-V-T (19 participants) and A-T-V (14 participants).

All stimuli were presented in a random block design using Experiment Creator X (Thompson & Kosarev, 2000), a freeware software program that allows experimenters to present media files using a response-prompting interface. After each test trial, participants made their responses using a single row of keys on the MacBook. Keys 1-9 represented semitone ranges 1-9, respectively, while the keys "0" and "–" represented semitone ranges of 10 and 11, respectively. These latter two keys were labelled appropriately. The space bar was used to proceed between trials once a response had been made.

Procedure. Participants entered the laboratory and were told that the experiment was designed to look at how different modal stimuli could contribute to judgments of musical information. Each participant received an information sheet and consent form that described further experiment details, risks, and benefits. Participants were also informed about the meaning of semitones and interval size. Upon receiving consent and after addressing any questions or concerns, participants were seated in the sound attenuating chamber. Prior to beginning any condition, participants were given examples of the smallest and largest semitone ranges possible (0 and 12 semitones, respectively) from both singers, creating a total of 4 practice trials. After the practice trials were completed, participants were instructed as to how to make responses during the testing phase. On each trial, participants were asked to estimate the semitone range they had just heard, seen, and/or felt, but base

their judgments on the auditory component alone. Participants were given a semitone range from 1-11 to respond with, thus allowing a total of 12 different responses. Participants were told that the message, "Hurry! Make your response now!" would display immediately after each trial had finished playing and that upon seeing this message, they had a maximum of 3 seconds to make a response. This restriction on response times was expected to ensure musically trained listeners responded primarily on relative pitch as a musically untrained listener would as oppose to referring to some categorical memory representation of what the specific interval sounds like (Russo & Thompson, 2005). To prevent issues with prior exposure and familiarity, only trials with semitone ranges from 1-11 were used in the testing phase. Prior to any bimodal condition testing, participants were instructed to base their judgments on the auditory information alone. Small breaks were given between testing conditions.

Statistical Analyses. Accuracy and precision served as dependent variables. Order was used as a between-subjects independent variable. Within-subject independent variables consisted of condition, singer, repetition (of a unique trial), and interval. Musical expertise served as a covariate in all analyses. Unless otherwise reported, effects of gender and musical expertise were non-significant. Accuracy and precision scores are reported in semitones.

Results

The data from the experiment contained issues with normality of accuracy and precision scores, assessed using Kolmogorov-Smirnov tests. However, sample sizes were considered large enough to approach central tendencies and parametric analyses were used.

Accuracy. Accuracy scores were subjected to a 3x2x3x11x2 mixed ANOVA that

used condition (audio-alone, visual-alone, vibrotactile-alone), singer (1^{st} or 2^{nd}), repetition (1^{st}, 2^{nd}, 3^{rd}), and interval (1-11 semitone difference) as within-subject variables and order (A-V-T, A-T-V) as the between-subjects variable. Condition had a significant effect on accuracy, $F(2, 56) = 8.45$, $p = .001$, $\eta^2 = .23$, $MSE = 122.36$. Pairwise comparisons revealed significant differences between all 3 conditions. Audio-alone ($M = 2.30$) produced the best accuracy scores, visual-alone ($M = 2.60$) produced the second best accuracy scores, and vibrotactile-alone ($M = 3.02$) produced the worst accuracy scores. The covariate, music expertise, was not significantly related to accuracy scores, $F(1, 28) = 1.36$, $p = .254$, $\eta^2 = .05$, $MSE = 56.41$. Mean accuracy scores were significantly better than chance for audio-alone, $t(32) = 12.92$, $p < .001$, visual-alone, $t(30) = 10.15$, $p < .001$, and vibrotactile-alone trials, $t(32) = 5.21$, $p < .001$ (See Figure 2). Maulchy's test indicated that the assumption of sphericity was violated for interval, $\chi^2(54) = 224.09$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .280$. Interval size had a significant effect on accuracy scores, $F(2.797, 33.330) = 15.88$, $p < .001$, $\eta^2 = .36$, $MSE = 529.31$. A simple plot of interval size revealed that participants were best at estimating lower interval sizes than larger interval sizes.

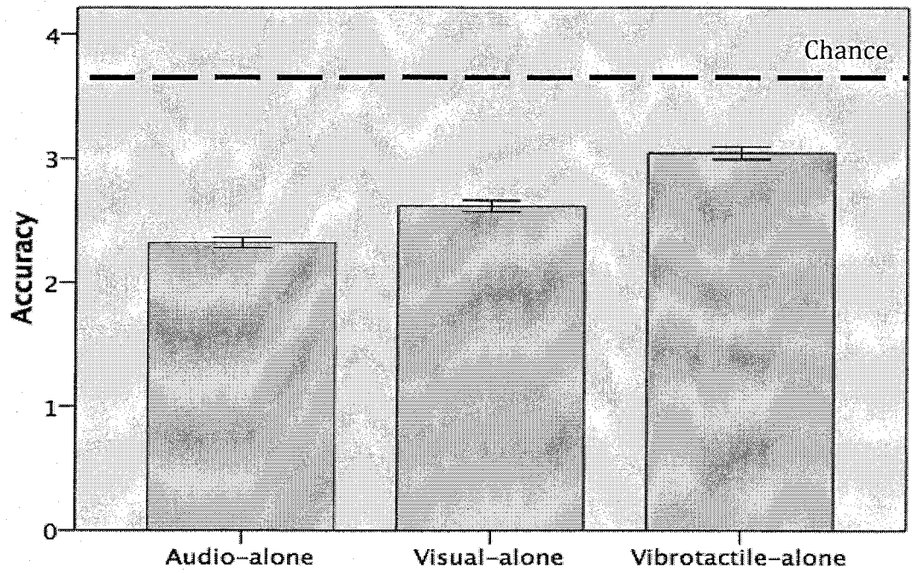

Figure 2. Mean accuracy scores for conditions in Experiment 1. Error bars represent +/- 1 standard error. The dashed line represents chance accuracy.

Singer type had a significant effect on accuracy, $F(1, 339) = 16.42$, $p < .001$, $\eta^2 = .05$, $MSE = 84.20$. Pairwise comparisons revealed that the second singer ($M = 2.52$) received significantly higher accuracy scores than the first singer ($M = 2.76$). A significant effect of order was also found, $F(1, 338) = 4.36$, $p = .037$, $\eta^2 = .01$, $MSE = 4.33$. Independent samples t-tests revealed that accuracy scores in the first order (A-V-T) were significantly better than accuracy scores in the second order (A-T-V) for audio-alone, $t(361) = 2.13$, $p = .034$, visual-alone, $t(316.374) = 2.017$, $p = .045$, and vibrotactile-alone trials, $t(361) = 3.06$, $p = .002$. Separate analyses of accuracy scores across conditions for each order were conducted. In both orders, audio-alone still produced significantly greater accuracy scores than visual-alone, which still produced significantly greater accuracy scores than vibrotactile-alone. No effect of repetition was found.

Precision. Precision scores were subjected to a 3x11x2 mixed ANOVA that used

condition (audio-alone, visual-alone, vibrotactile-alone) and interval (1-11 semitone difference) as within-subjects variables and order (A-V-T, A-T-V) as the between-subjects variable. Condition did not have a significant effect on precision, $F(2, 56) = 2.10$, $p = .132$, $\eta^2 = .07$, $MSE = 4.77$ (See Figure 3). No effect of order or music experience was found. Interval size had a significant effect on precision, $F(10, 280) = 4.45$, $p < .001$, $\eta^2 = .14$, $MSE = 2.40$. A simple plot did not reveal any trends of precision based on interval size.

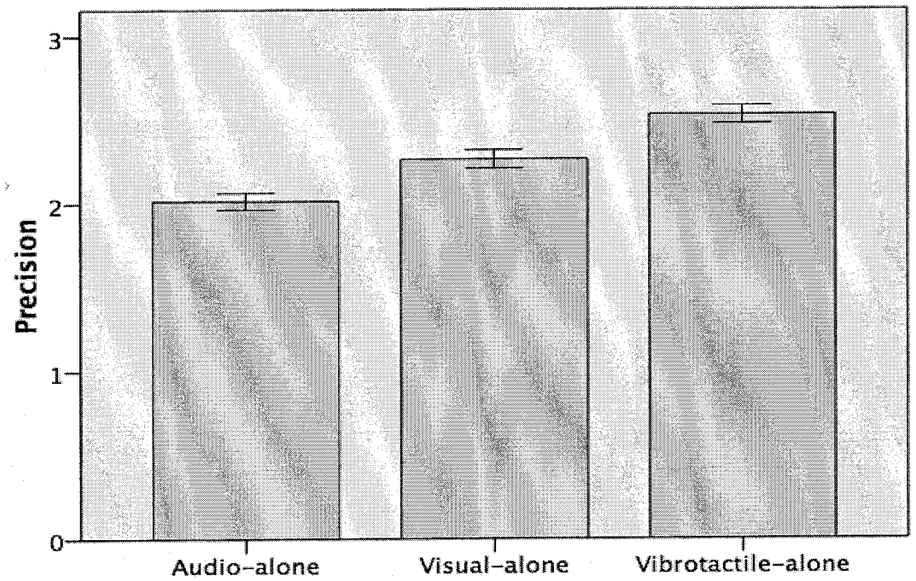

Figure 3. Mean precision scores for conditions in Experiment 1. Error bars represent +/- 1 standard error.

Discussion

As expected, the audio-alone condition received the best levels of accuracy, the visual-alone condition received the second best levels of accuracy, and the vibrotactile-alone condition received the worst levels of accuracy. Precision scores followed a similar pattern. A number of reasons can be given to support these findings. As previously mentioned, individuals with normal hearing do not have as much prior exposure to visual

and vibrotactile presentations of music compared to auditory presentations. From both a language and music perspective, visual information serves as a secondary source of information while auditory information serves as a primary source of information[1]. This explains why the auditory component of music can be isolated but not necessarily the visual component. Similarly, speech is difficult to understand from visual features alone (lip-reading), but easily understood from auditory features alone. Thus, it is possible that participants were not prepared or comfortable with making judgments of musical stimuli from isolated visual and vibrotactile signals. Indeed, a few participants expressed their doubts that any relevant information could be gained from isolated visual and vibrotactile signals. Yet, isolated visual components of music have been shown to still preserve the ordering of interval size, although interval size magnitude is not as effectively preserved as when isolated auditory components are used (Thompson, Graham, & Russo, 2005).

One concern is that the audio-alone condition was always presented first, followed by either visual-alone or vibrotactile-alone trials. Thus, the audio-alone condition may have produced superior results simply due to primacy. However, a recency effect could have also occurred, but did not (accuracy trends remained the same regardless or condition order). Additionally, participants had very little practice when they started the audio-alone condition, as oppose to the visual-alone and vibrotactile-alone conditions. It is more likely that the superior accuracy scores in the audio-alone condition are due to extensive prior exposure to music through the auditory modality.

Despite these concerns, participants still performed significantly better than chance in all 3 conditions, suggesting that there was information, inherent in both the visual and

[1] This argument, however, cannot be applied to individuals that are deaf and hard-of-hearing, since these individuals rely on the visual and tactile components of speech and music for understanding.

vibrotactile signals, that was of benefit towards estimations of interval size. This finding is consistent with the hypothesis that visual and vibrotactile information can enhance musical judgments when paired with audio. Such a benefit should theoretically extend to bimodal presentations.

Experiment 2: Bimodal Presentations

Experiment 1 revealed that relatively accurate estimates of interval size can be made using auditory, visual, and vibrotactile signals. The next experiment was designed to investigate whether bimodal presentations could provide even greater benefits for musical judgments. In addition, the following experiment examined whether crossmodal integration was dependent on learning by comparing estimates of interval size for audio-visual (relatively learned) and audio-vibrotactile (relatively unlearned) trials.

Based on the hypothesis that crossmodal integration is not necessarily dependent on learning or familiarity, I predicted the auditory-visual and auditory-vibrotactile conditions would produce similar levels of accuracy. Both these conditions were also expected to produce significantly greater levels of accuracy than auditory-alone.

Methods

A total of 40 Ryerson undergraduate students (average age = 20.1 years, 2 male, 4 left-handed) with an average of 3.1 years of musical training (SD = 4.15) participated for course credit. These participants were different from the participants used for Experiment 1. The test stimuli, design, procedure, and statistical analyses were similar to Experiment 1 except for the following modifications. The experiment consisted of 3 conditions: audio-alone (A), audio-visual (AV), and audio-vibrotactile (AT). In the audio-alone condition, participants only heard the intervals being sung. In the audio-visual condition, participants heard and saw the intervals being sung. In the audio-vibrotactile condition, participants heard and felt the intervals being sung. In order to prevent a ceiling effect from occurring in terms of accuracy, and to provide greater opportunity for crossmodal integration to occur (inverse effectiveness rule), the auditory component in each condition was masked over by

white noise to obtain a signal-to-noise ratio of -6dB, which rendered the intervals barely audible. Each condition consisted of 44 trials (22 unique trials repeated twice for 3 different conditions for a total of 132 trials) with an equal number of trials allocated to each singer and interval. The conditions were counterbalanced across participants to create four different orders, with 10 participants in each order: A-AV-AT, A-AT-AV, AV-AT-A, and AT-AV-A.

Results

The data from the experiment contained issues with normality of accuracy and precision scores, assessed using Kolmogorov-Smirnov tests. However, sample sizes were considered large enough to approach central tendencies and parametric analyses were used.

Accuracy. Accuracy scores were subjected to a 3x2x3x11x4 mixed ANOVA with condition (audio-alone, visual-alone, tactile-alone), singer (1^{st} or 2^{nd}), repetition (1^{st}, 2^{nd}, 3^{rd}), and interval (1-11 semitone difference) as within-subjects variables and order (A-AV-AT, A-AT-AV, AV-AT-A, AT-AV-A) as the between-subjects variable. Maulchy's test indicated that the assumption of sphericity was violated for condition, $\chi^2(2) = 7.11$, $p = .029$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .841$. The effect of condition was significant, $F(1.682, 58.884) = 9.96$, $p < .001$, $\eta^2 = .22$, $MSE = 112.80$. Pairwise comparisons revealed that accuracy for audio-visual trials (M = 2.55) was significantly greater than accuracy for audio-vibrotactile trials (M = 3.00). Differences in accuracy between audio-visual (M = 2.55) and audio-alone (M = 2.75) conditions approached marginal significance. No significant difference was found between audio-alone (M = 2.75) and audio-vibrotactile (M = 3.00) conditions (see Figure 4). No effect of repetition was found.

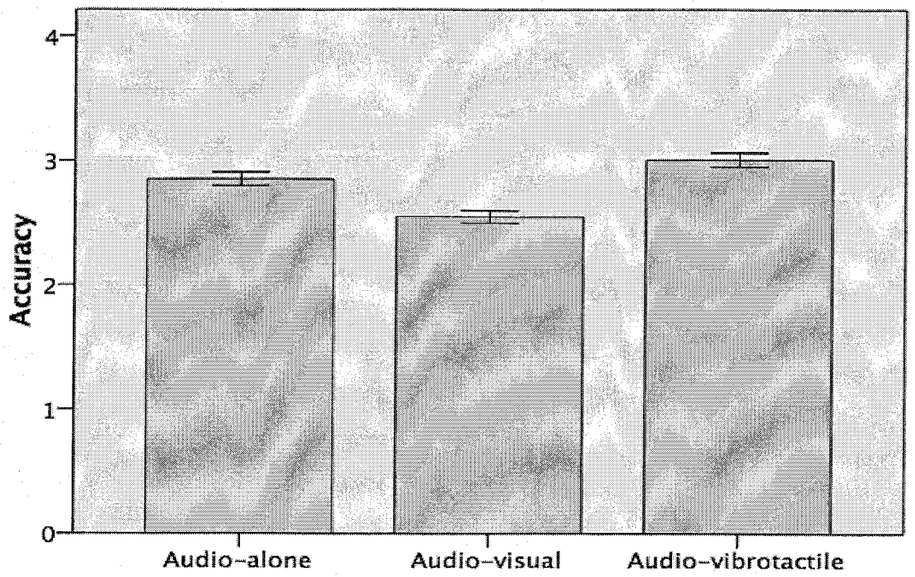

Figure 4. Mean accuracy scores for conditions in Experiment 2. Error bars represent +/- 1 standard error.

The effect of singer was marginally significant, $F(1, 35) = 3.95$, $p = .055$, $\eta^2 = .10$, $MSE = 47.65$. Order alone had no significant effect on overall accuracy scores, but a significant interaction between order and condition was found, $F(6, 70) = 3.15$, $p = .009$, $\eta^2 = .21$, $MSE = 30.05$. Contrasts revealed that order had a significant effect on accuracy between audio-alone and audio-vibrotactile conditions and between audio-visual and audio-vibrotactile conditions. Profile plots revealed that audio-vibrotactile accuracy scores were worst when the audio-vibrotactile condition was presented first (order AT-AV-A) (see Figure 5). Maulchy's test indicated that the assumption of sphericity was violated for interval, $\chi^2(54) = 212.74$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .339$. Interval had a significant effect on accuracy, $F(3.394, 118.793) = 13.03$, $p < .001$, $\eta^2 = .27$, $MSE = 382.82$. A simple plot revealed that participants were poorer at accuracy for intervals that were larger in size.

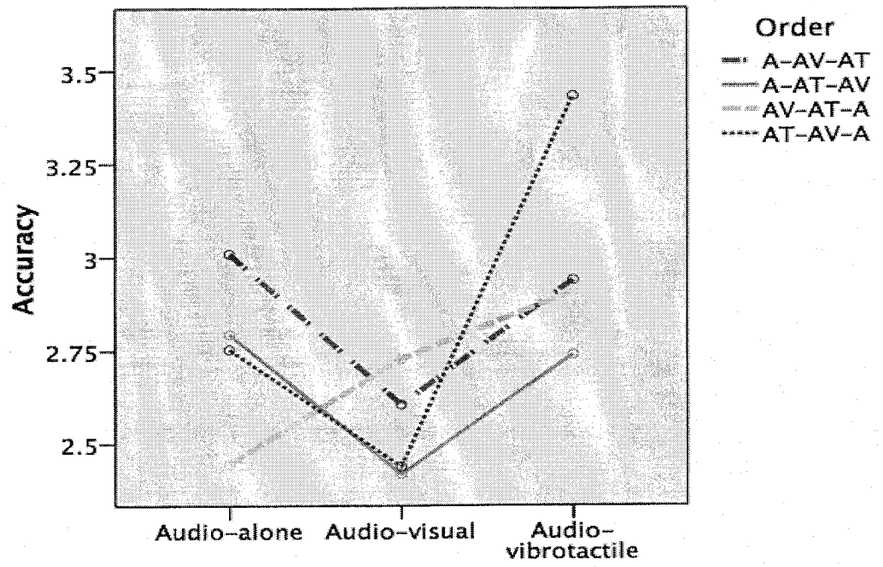

Figure 5. Plot of accuracy scores for each condition based on order for Experiment 2.

The covariate, music expertise, was significantly related to accuracy scores, $F(1, 35) = 4.56$, $p = .04$, $\eta^2 = .12$, $MSE = 181.47$. Planned contrasts revealed that higher levels of music experience significantly improved accuracy scores for audio-vibrotactile trials, $t(36) = 2.49$, $p = .018$.

Precision. Precision was subjected to a 3x11x4 mixed ANOVA with condition (audio-alone, audio-visual, audio-vibrotactile) and interval (1-11 semitone difference) as the within-subjects variables and order (A-AV-AT, A-AT-AV, AV-AT-A, AT-AV-A) as the between-subjects variable. Condition had a significant effect on precision scores, $F(2, 70) = 7.48$, $p = .001$, $\eta^2 = .18$, $MSE = 14.75$. Pairwise comparisons revealed that precision scores were significantly greater in the audio-alone condition ($M = 2.01$) than in the audio-vibrotactile ($M = 2.34$) condition. No significance differences in precision were found between the audio-visual and audio-vibrotactile groups or audio-alone and audio-visual groups (See Figure 6).

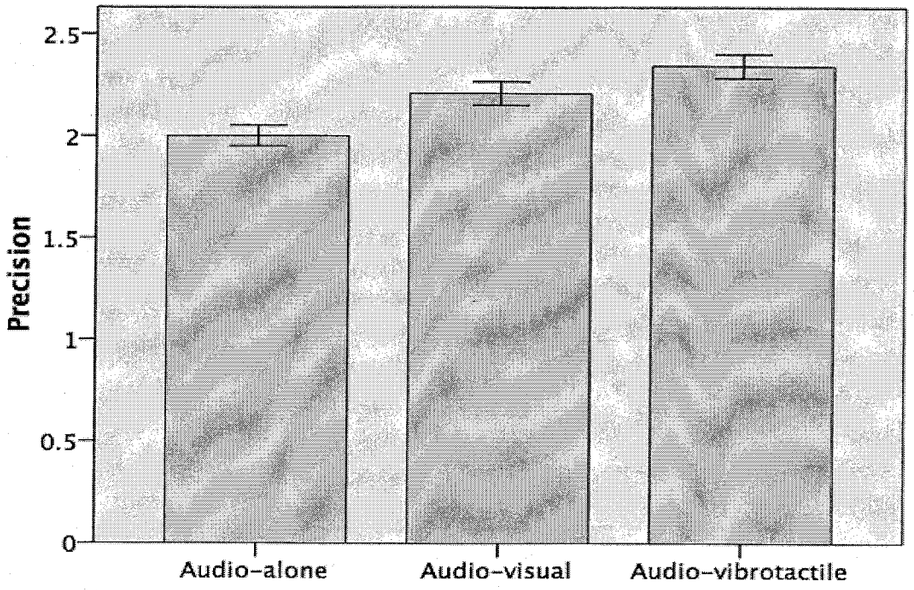

Figure 6. Mean precision scores for conditions in Experiment 2. Error bars represent +/- 1 standard error.

The covariate of music experience had no significant effect on precision scores, $F(1, 35) = 0.93$, $p = .341$, $\eta^2 = .03$, $MSE = 9.41$. Maulchy's test indicated that the assumption of sphericity was violated for interval, $\chi^2(54) = 92.57$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .605$. A marginally significant effect of interval was found, $F(6.049, 211.371) = 2.00$, $p = .066$, $\eta^2 = .05$, $MSE = 3.97$.

Discussion

Results do not support the hypothesis that auditory judgments of musical stimuli can benefit from vibrotactile information, at least not in the absence of training. In fact, accuracy was slightly worse on audio-vibrotactile trials than audio-alone trials, despite the fact that the audio signal was distorted with white noise. Participants may have been distracted by the vibrotactile signals, as they were given no instructions on how to use or

interpret it for the instructed task. The vibrotactile source could have been perceived as artificial relative to how participants normally experience vibrations from music (i.e., speakers, furniture, instruments).

Despite this finding, accuracy was significantly improved with visual information, suggesting that watching the intervals being sung was of benefit to participants' estimates of interval size. The visual signals consisted of facial expressions and physical gestures (eye and mouth widening, head and eyebrow displacement) that likely served as distinctive illustrators for each interval. Extensive exposure and familiarity with visual music features may have given participants a better understanding of how to employ the visual signal to their advantage, despite being given no instructions on how to use or interpret it for the instructed task. Alternatively, the audio-visual signal may be more beneficial than the audio-vibrotactile signal because the visual signal provides complimentary information while the vibrotactile signal does not. Instead, the vibrotactile signal may only provide redundant information since it is driven by the same stimulus (i.e., pressure waves) as the auditory signal.

Precision scores were best for the audio-alone condition, suggesting that participants were more consistent and confident in their responses when auditory information was presented alone. The opposite was true for the audio-vibrotactile condition, possibly due to similar reasons given for this condition's poor accuracy scores. Results suggest that the participants were not as certain of their responses for the intervals in the audio-vibrotactile condition, possibly because the vibrotactile signal distracted participants from listening to the auditory signal.

A comparison of experiments 1 and 2 revealed that unimodal signals received similar accuracy scores as bimodal signals. This was unexpected as the bimodal signals were expected to receive accuracy scores that were superior to the scores acquired from any of the unimodal signals. Even if participants are not using crossmodal integration when perceiving the bimodal signals, these signals nevertheless contain more information than the unimodal signals and should give the participant a better understanding of what he or she perceived.

The following reason can be given as to why this result occurred. Separate groups of participants were used in experiments 1 and 2, with the latter group reporting a lower level of music expertise (3.1 years versus 4.1 years in Experiment 1). In addition, music expertise was found to be a significant covariate for accuracy scores in both bimodal conditions. Thus, music expertise may be not only beneficial, but also necessary to benefit from crossmodal signals in a musical context. Indeed, greater benefits from the crossmodal signals may have been found had the participants from Experiment 2 been restricted to musicians.

Music experience significantly improved accuracy scores for audio-vibrotactile trials. This is not surprising, as formal music training consists of extensive exposure and feedback from visual and tactile modalities in addition to the auditory modality. This suggests that training may be necessary in order to benefit from crossmodal presentations of music. The extent of training necessary, however, before crossmodal integration benefits can be fully made remains undetermined.

Experiment 3: Vibrotactile Training

Participants did not gain significant musical judgment benefits from audio-vibrotactile signals compared to audio-alone. One possible reason for this is that humans have minimal experience using vibrotactile signals within a music judgment framework. The third experiment served as a follow-up to the second experiment and looked at whether minimal vibrotactile training and audio-vibrotactile mapping could improve the use of vibrotactile information towards estimates of interval size.

The experiment was divided into 2 separate parts. Part 1 was designed to investigate whether vibrotactile training alone could improve judgments of audio-vibrotactile musical stimuli. Part 2 was designed to investigate whether the mapping of vibrotactile information onto audio information was necessary for improving judgments of audio-vibrotactile musical stimuli. For Part 1, I predicted that participants would significantly improve their accuracy during training. I also predicted the auditory-visual and auditory-vibrotactile conditions would receive similar levels of accuracy to one-another and greater levels of accuracy than audio-alone. Similar predictions were made for Part 2.

Methods

A total of 3 undergraduate students (average age = 22.7 years, 2 male, all right-handed) with an average 15.7 years of musical training (SD = 14.8) volunteered their time to participate in this experiment. The test stimuli, design, procedure, and statistical analyses were similar to Experiment 1 except for the following modifications. Each participant was exposed to 4 training sessions (2 blocks of 44 trials in each session for a total of 8 blocks) and 5 testing sessions over 4 separate days, with all days occurring within 3 weeks from start to finish. Participants were trained using a set of stimuli that were created using a

synthesized voice ('Choir Aahs' sound file, Vocal Writer 2.0.1). This allowed participants to train their vibrotactile judgments of the semitone ranges but prevented them from gaining experience with the actual test stimuli (real female voices). During training sessions, participants were presented with a trial and then asked to verbalize a response to the experimenter. Upon responding with an estimate, the actual interval size was presented to the participant on the computer screen. Small breaks were given between training and testing blocks.

As previously mentioned, the experiment was split into 2 parts. Part 1 was designed to investigate whether training with vibrotactile-alone stimuli would improve judgments of audio-vibrotactile musical stimuli; it consisted of 4 testing sessions and 6 blocks of vibrotactile-alone training. The first testing session served as a baseline for all conditions; participants were presented with audio-alone, audio-visual, and audio-vibrotactile conditions (similar to Experiment 2), counterbalanced across participants. The next 3 testing sessions only included the audio-vibrotactile condition. During training, participants were deafened using earplugs and sound-attenuating headphones.

Part 2 was designed to investigate whether the mapping of vibrotactile information onto audio information would provide an even greater benefit towards judging audio-vibrotactile musical stimuli; it consisted of one final testing session for audio-vibrotactile stimuli preceded by two blocks of audio-vibrotactile training. A concern was that participants would rely solely on the auditory information and disregard the vibrotactile information. Thus, hearing was reduced slightly using sound-attenuating headphones. This allowed participants to perceive auditory information from the Tactaid skin transducers but at a barely audible level.

Table 2.
Testing and training for each participant in Experiment 3.

Participant	Baseline	Train	Test 2	Train	Test 3	Train	Test 4	Train	Test 5
1	A AV AT	T	AT	T	AT	T	AT	AT	AT
2	AV AT A	T	AT	T	AT	T	AT	AT	AT
3	AT A AV	T	AT	T	AT	T	AT	AT	AT

Results

The data from the experiment contained issues with normality of accuracy and precision scores, assessed using Kolmogorov-Smirnov tests. However, the data was considered large enough to approach central tendencies and parametric analyses were used.

Training. Accuracy between training blocks 1 and 8 was subjected to a paired-samples t-test. Participants significantly improved in accuracy from blocks 1 to 8, $t(131) = 7.43$, $p < .001$ (See Figure 7). Participants also reported that the estimates became easier over blocks of training.

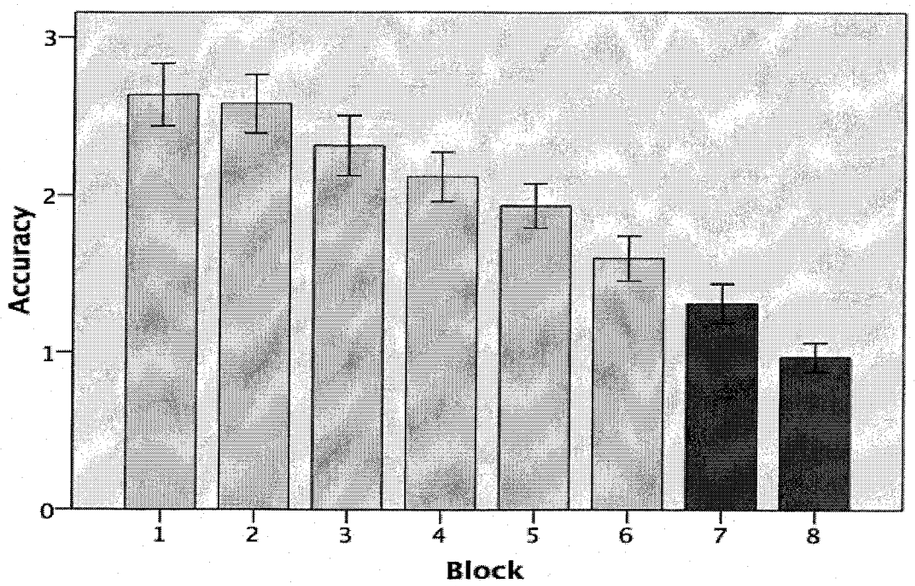

Figure 7. Mean accuracy scores across blocks of training for Experiment 3. The first 6 blocks (1-6) consisted of vibrotactile-alone training, while the last 2 blocks (7-8) consisted of audio-vibrotactile training. Error bars represent +/- 1 standard error.

Accuracy. Baseline accuracy scores were subjected to a 3x2x11x3 mixed ANOVA with condition (audio-alone, audio-visual, audio-vibrotactile), singer (1^{st}, 2^{nd}) repetition (1^{st}, 2^{nd}, 3^{rd}), and interval (1-11 semitone difference) serving as within-subjects variables. A significant condition effect was not found, $F(2, 20) = 2.17$, $p = .140$, $\eta^2 = .18$, $MSE = 4.25$ (see Figure 8). No of repetition was found. Singer had a significant effect on accuracy, $F(\ ) = 26.89$, $p < .001$, $\eta^2 = .73$, $MSE = 353.94$. Interval also had a significant effect on accuracy, $F(1, 10) = 32.05$, $p < .001$, $\eta^2 = .76$, $MSE = 110.20$. A simple plot revealed that participants were quite good at judging small interval sizes but were poorer at judging larger interval sizes.

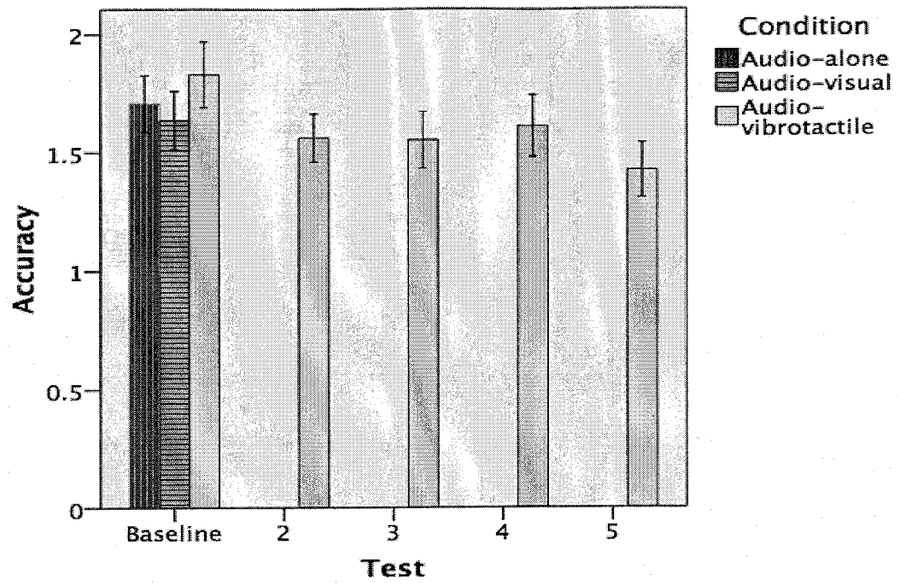

Figure 8. Mean accuracy scores across tests in Experiment 3. Error bars represent +/- 1 standard error.

Paired samples t-tests were performed to determine if vibrotactile training (blocks 1-6) significantly improved audio-vibrotactile accuracy scores over audio-alone and audio-visual trials. Test 4 audio-vibrotactile accuracy scores were compared to baseline audio-alone and audio-visual accuracy scores. No significant differences in accuracy were found between audio-vibrotactile trials on Test 4 and baseline audio-alone trials, $t(32) = 0.20$, $p = .847$. No significant differences in accuracy were found between audio-vibrotactile trials on Test 4 and baseline audio-visual trials, $t(22) = 0.31$, $p = .759$.

Paired samples t-tests were performed to determine if audio-vibrotactile training (blocks 7-8) significantly improved audio-vibrotactile accuracy scores over audio-alone and audio-visual trials. Test 5 audio-vibrotactile accuracy scores were compared to baseline audio-alone and audio-visual accuracy scores. Test 5 audio-vibrotactile trials received significantly better accuracy scores compared to audio-alone trials, $t(32) = 3.04$, $p = .005$.

Audio-vibrotactile trials also received significantly better accuracy scores compared to audio-visual trials, $t(22) = 2.23$, $p = .036$.

Precision. Baseline precision scores were subjected to a 3x11 mixed ANOVA with condition (audio-alone, audio-visual, audio-vibrotactile) and interval (1-11 semitone difference) serving as the within-subjects variable. No significant differences in precision were found between conditions, $F(2, 22) = 0.25$, $p = .784$, $\eta^2 = .02$, $MSE = 0.13$ (See Figure 9). . Maulchy's test indicated that the assumption of sphericity was violated for interval, $\chi^2(54) = 100.82$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .323$. Interval had a significant effect on precision, $F(3.227, 35.492) = 21.29$, $\eta^2 = .66$, $MSE = 48.78$. A simple plot revealed that participants were quite consistent when judging small interval sizes but were less consistent when judging larger interval sizes.

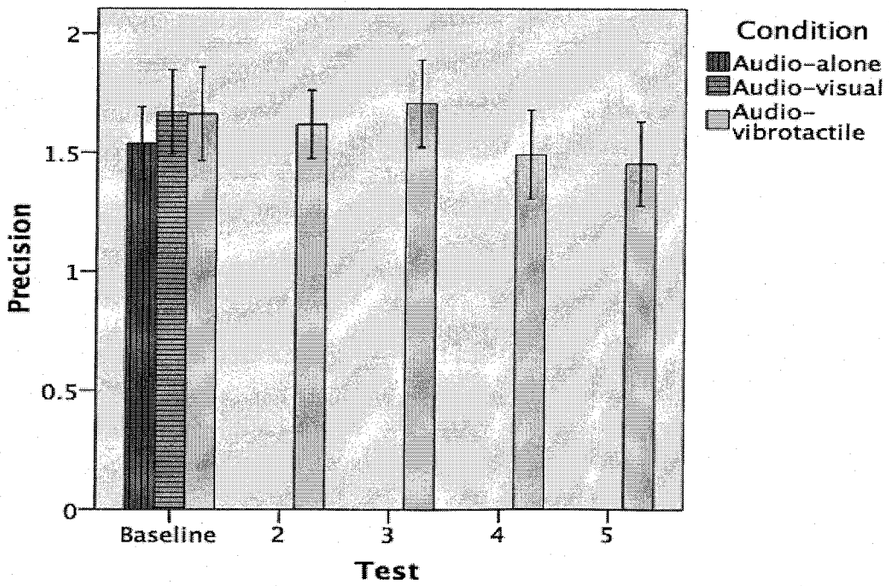

Figure 9. Mean precision scores across tests in Experiment 3. Error bars represent +/- 1 standard error.

Paired samples t-tests were performed to determine if vibrotactile training (blocks 1-6) significantly improved audio-vibrotactile precision scores over audio-alone and audio-visual trials. No significant differences were found in precision between audio-vibrotactile trials on Test 4 and baseline audio-alone trials, $t(32) = 0.28$, $p = .777$. Similarly, no significant differences were found in precision between audio-vibrotactile trials on Test 4 and baseline audio-visual trials, $t(32) = 0.64$, $p = .526$.

Paired samples t-tests were performed to determine if audio-vibrotactile training (blocks 7-8) significantly improved audio-vibrotactile precision scores over audio-alone and audio-visual trials. No significant differences were found in precision between audio-vibrotactile trials on Test 5 and baseline audio-alone trials, $t(32) = 1.01$, $p = .321$. Similarly, no significant differences were found in precision between audio-vibrotactile trials on Test 5 and baseline audio-visual trials, $t(32) = 1.33$, $p = .192$.

Discussion

Results support the hypothesis that audio-vibrotactile training can improve the benefits gained from the vibrotactile signal in the audio-vibrotactile condition. If lack of familiarity and experience with the vibrotactile signal do in fact explain findings from experiments 1 and 2, it appears that the training paradigm somewhat relieved these issues. The subsequent mapping between vibrotactile and auditory information significantly improved the use of vibrotactile signals towards estimates of interval size compared to the baseline accuracy scores from the audio-alone and audio-visual conditions.

As can be seen from Figure 8, the audio-vibrotactile condition started off with the worst accuracy scores. However, accuracy improved in the audio-vibrotactile condition as training progressed until it was producing better judgments of interval size than the audio-

alone condition and even the audio-visual condition. Precision scores were also similar across conditions. Together, these findings suggest that participants did gain significant benefit from the vibrotactile features. One explanation why greater benefits from the vibrotactile modality were not found is that participants already had extensive experience with interval size in the auditory modality. Indeed, 2 of the 3 participants already had high levels of musical expertise (17.5 and 29.5 years). These participants may have found the auditory information more than sufficient to produce estimates of interval size, thereby minimizing the utility of visual or vibrotactile signals relative to another participant that had little or no music experience. Another possibility is that further crossmodal mapping of auditory and vibrotactile modalities could have generated greater benefits from the vibrotactile signal towards estimates of interval size.

An observation worth mentioning is the motivational factor inherent in this experiment. Participants were motivated to improve upon each training block and test due to feedback and performance updates. This design was meant to generalize the findings somewhat towards real-life music experiences, where motivational factors are usually present (e.g., understanding melodic and harmonic patterns for entertainment, emotional, social, and/or educational value). Motivation and high levels of musical expertise explain why accuracy and precision scores were better than scores from participants in experiments 1 and 2.

Findings from this study, however, should be interpreted with caution as the small sample size restricts generalization. Perhaps a more extensive connection between audio and vibrotactile information (e.g., a larger vibrating contactor) is necessary in order for participants to have truly benefitted from the vibrotactile signal. Nevertheless, this

experiment provides promising evidence that individuals can benefit from an auditory-vibrotactile music signal with minimal amounts of training.

General Discussion

Overall results suggest that benefits from crossmodal integration within a musical context are to a certain extent dependent on prior exposure and familiarity with the crossmodal signal. For bimodal presentations, the vibrotactile signal failed to provide significant benefits towards accuracy or precision for estimates of interval size. However, estimates of interval size were still better than chance when the participant could only utilize vibrotactile information. This is suggestive of a significant judgment benefit from the vibrotactile signal, which nonetheless appears irrelevant when auditory information is present. Visual information also provides judgment benefits, but unlike vibrotactile signals, this information is utilized to significantly improve auditory estimates of interval size.

The results suggest that a learning component exists for crossmodal integration in music perception. With minimal amounts of training, estimates significantly benefitted from the vibrotactile signal. Two main hypotheses can be given as to why audio-vibrotactile benefits did not occur without training or why greater audio-vibrotactile benefits were not found. The first hypothesis is that the vibrotactile signal provided little or no novel information to the participant. In other words, the vibrotactile signal may have just provided superfluous details about the intervals. This theory may be supported if we assume that individuals cannot interpret frequency information inherent in vibrotactile signals. However, humans do have some ability to make discriminations on the basis of vibrotactile frequency (Pongrac, 2008; Rothenberg, Verillo, Zahorian, Brachman, & Bolanowski, 2006), but see Appendix D for further discussion. The vibrotactile signals also provide onset and offset cues, informing the participant of tempo and note durations. Indeed, some

participants in Experiment 2 (bimodal presentations) reported that the Tactaids helped to determine when the auditory information was presented for each trial.

The second hypothesis is that participants were unable to utilize the vibrotactile information due to issues with exposure and/or familiarity within a musical context. In other words, relevant information is present in the vibrotactile signal but participants were either unaware of such information or did not understand how to use it effectively. Although it is common to experience vibrations from music (e.g., bass), such exposures, as previously mentioned, are from the low-end of the frequency spectrum. Exposure to mappings of all vibrotactile frequencies onto auditory frequencies is therefore uncommon (or at least not as common as the mapping of visual and auditory music information). Because of this, individuals never really learn to make musical associations between what they are hearing and what they are feeling, despite the fact that both modalities stem from the same energy source. This hypothesis suggests that crossmodal integration in music requires learning before any potential benefits can be received.

Although previous speech studies have shown immediate benefits from using tactile information (Fowler & Dekle, 1991; Gick, Jóhannsdóttir, Gibraiel, & Mühlbauer, 2008), these benefits may be limited to variable onset/offset cues and other temporal cues associated with each phoneme rather than with frequency information. However, temporal dynamics are far more critical for speech than for music (Phillips & Farmer, 1990; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Tallal, Miller, & Fitch, 1993). Since note durations tend to be significantly longer than speech phonemes (Fraisse, 1974), the need for sensitivity in music to gauge temporal structure is lower than in speech. Instead, music perception relies on pitch variations with fine temporal structure, a common component of

all musical systems for the production of melodies (Attneave & Olson, 1971; Warrier & Zatorre, 2002). Another point of comparison between the present study and tactile speech research is the use of the Tadoma Method (i.e., a human speaker was present) versus the use of an electronic medium (Tactaids). Indeed, the Tactaid skin transducers may not have been able to provide as lawful or natural an association between the auditory and tactile signals as the Tadoma Method can.

Within experiments 1 and 2, singer type had a significant effect on accuracy and precision scores. One reason for this is that the second singer showed more facial expressions and movements than the first singer. Another reason is that the second singer sung her notes more succinctly than the first singer. In other words, the second singer provided better cues for the onset and offset of each note (i.e., sharper attack, greater head movement during note production). This allowed for better comparison between the notes, especially when white noise was masking the auditory signal.

Although average pitch height was equated for the intervals, one potential concern was that participants might base their judgments of pitch on the first note alone and disregard the second note. First, participants were not told about the distribution of notes (i.e., if the first note is lower, this means that the interval will be larger), making it difficult for participants to judge interval size without paying attention to the second note. Second, individuals are adept at using relative pitch but are poor or unable to process pitch in absolute terms (Brown, Sachs, Cammuso, & Folstein, 2002). This makes it difficult for participants to judge the exact value of an isolated note, especially when other notes of proximal spectral sensitivity are equally likely. However, this concern warrants further study.

Limitations and Future Directions

There were a number of limitations from the experiments, some of which have already been discussed but will be reiterated here. The first limitation was that participants might have found the Tactaids aversive. Although no participant reported any pain or unwanted sensations from the Tactaids, holding the skin transducers may have seemed trivial to the participant in terms of the experimental task. This, coupled with a lack of familiarity with higher vibrotactile music frequencies, possibly restricted any potential benefits that could have been gained from the vibrotactile signal. Giving more thorough instructions on how the Tactaids worked and what kind of musical information the vibrotactile signal could provide could have reduced this limitation.

A second limitation was that the auditory component in the bimodal testing conditions was not sufficiently masked. Although most participants had difficulty at first identifying the 2 semitones on each trial over the white noise, identification likely became easier over time. Participants may have been convinced that the audio signal was sufficient to provide good estimates of interval size. Alternatively, masking of the auditory signal may have been too extensive, requiring most or all of a participant's attentional resources. This would have limited the attentional resources available for the visual or vibrotactile signal.

Another limitation was that participants might not have been motivated to provide their most capable estimates of interval size. Participants received no feedback during testing in experiments 1 and 2, but did receive feedback in experiment 3. Significantly greater accuracy and precision scores were produced in experiment 3 compared to the first two experiments. Although such a difference could be attributed to higher levels of musical expertise, participants in experiment 3 were nonetheless motivated to perform better.

The following research can be used towards a number of future directions. The experiments could be repeated taking note of the aforementioned limitations and removing or reducing them if possible. For example, the experiments could be replicated using a live singer and the Tadoma Method as a vibrotactile source. A follow-up to the bimodal experiment would be to vary the level of white noise within each signal on a trial-to-trial basis to see if dependence on the accompanying signal (visual or vibrotactile) grows as the auditory signal is gradually distorted. The inverse effectiveness rule (Stein, Meredith & Wallace, 1994) would predict greater benefits from crossmodal integration as the auditory signal was diminished. However, as the distortion of the auditory signal increases, a point may be reached where the crossmodal signal fails to provide as good an estimate of interval size as the auditory signal alone (without distortion). This follow-up study could have implications towards the use of visual and vibrotactile signals as a sensory augmentation aid or sensory substitution aid to music perception for those that have diminished or no hearing, respectively.

Another potential follow-up study could look at whether extensive training could improve the use of vibrotactile signals towards musical judgments. Two groups of participants would be used: professional musicians with tactile feedback (e.g., violinists) and professional musicians with no tactile feedback (e.g., pianists). If vibrotactile benefits are based on prior experience, then musicians with extensive tactile feedback should make significantly better estimates of interval size than musicians with no tactile feedback.

Finally, the experiments could be replicated using a completely between-subjects design. This would eliminate any concerns with order (primacy and recency effects, fatigue,

novelty). New findings could potentially be discovered using this design that were not found using a within-subjects design.

These future directions merit continued research interest in the use of crossmodal signals as an aid to music perception. Vibrotactile signals are of significant interest in a variety of musically related areas, including performance, entertainment, speech and music discrimination, and perception of emotion. Understanding the potential benefits of and prerequisites for using vibrotactile signals will therefore likely become a more popular area of research.

Conclusions

To my knowledge, this is the first study to investigate potential benefits of a vibrotactile signal towards quantitative judgments of music. It is also the first study to demonstrate benefits of visual and vibrotactile information supporting accuracy and precision of interval size judgments. Although no significant advantages were gained from pairing vibrotactile information with auditory information for musical stimuli without prior learning, I anticipate advantages to be found with training and familiarization. Humans have a remarkable ability to integrate congruent aspects of time-varying and frequency-varying information obtained from different modalities in order to disambiguate stimuli of interest such as music. Future research should further investigate the manner and extent to which vibrotactile information can support the musical experience.

Appendix A

Crossmodal integration relies on a certain level of temporal congruency between modal signals. Prior to completing Experiment 2, an additional experiment was started that used incongruent presentations of audio-visual and audio-vibrotactile signals. The purpose of this experiment was to assess whether participants were integrating the modal signals (audio and visual, audio and vibrotactile) as oppose to simply averaging a response from the two signals. If individuals are sensitive to the time-varying aspects of music, then crossmodal integration should be reduced and accuracy should consequently decrease as latency increases between the modal signals. Previous speech research has shown that crossmodal integration is reduced when sufficient latency (on the magnitude of seconds) is introduced between the auditory and visual signals (Jones & Jarick, 2006). A similar effect could potentially be found using crossmodal music stimuli.

A group of 16 participants (20.6 years, 6 male, all right-handed) with an average of 3.5 years of musical training (SD = 5.3) were used in this experiment. The test stimuli, design, procedure, and statistical analyses were similar to Experiment 2 except for the following modifications. The bimodal signals (audio-visual and audio-vibrotactile) were separated in time by varying stimulus onset asynchronies, with presentation of the auditory signal synchronous or prior to presentation of the second modal signal (visual or vibrotactile). Trials were created in which the visual and vibrotactile signals lagged the auditory signal by 0, 200, 400, 600, 800, 1000, 1200, or 1400ms. The temporal constraints in the McGurk effect suggest fusion is more likely with auditory lag rather than lead (Wassenhove, Grant, & Poeppel, 2006). To ensure an appropriate number of exposures to each magnitude of latency but prevent the conditions from lasting too long, a select range of

intervals (1, 3, 5, 7, 9, and 11 semitones) were used. This selection reduced the number of trials required within each bimodal condition but maintained a consistent spread of interval size. Thus, participants were presented 96 trials (6 intervals x 2 singers x 8 latencies) each in the audio-visual and audio-vibrotactile conditions. No auditory-alone condition was used. Small breaks were given between test conditions.

Results. Accuracy was assessed using a 2x8x2x6x2 mixed ANOVA with condition (audio-visual, audio-vibrotactile), latency (0, 200, 400, 600, 800, 1000, 1200, 1400ms), singer (1^{st}, 2^{nd}), repetition (1^{st}, 2^{nd}), and interval (1,3,5,7,9,11 semitone difference) serving as within-subjects variables. Order (AV-AT, AT-AV) served as a between-subjects variable. No effects were found for condition, $F(1, 3) = 0.374$, $p = .584$, $\eta^2 = .11$, $MSE = 1.02$, or latency, $F(7, 21) = 0.84$, $p = .569$, $\eta^2 = .22$, $MSE = 2.17$. Singer had a significant effect on accuracy scores, $F(1, 3) = 11.36$, $p = .043$, $\eta^2 = .79$, $MSE = 3.35$. The covariate of music experience had no significant effect on accuracy scores, $F(4, 3) = 1.42$, $p = .403$, $\eta^2 = .65$, $MSE = 68.76$. No significant effect of order was found.

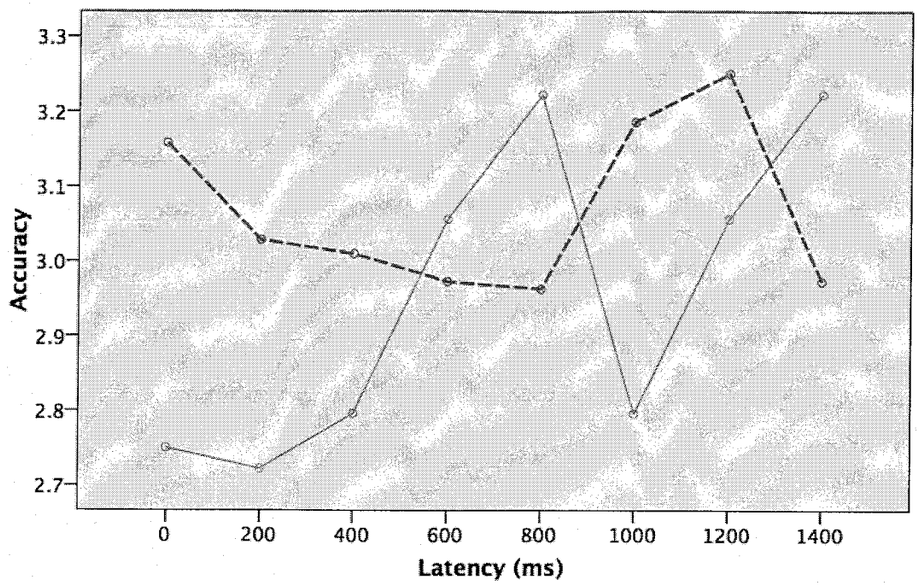

Figure 10. Mean accuracy scores across latencies for incongruent bimodal presentations. The dashed line represents the audio-visual condition and the solid line represents the audio-vibrotactile condition.

Discussion. The findings do not support the hypothesis that crossmodal integration is in fact occurring between the bimodal signals. Instead, participants may have been making estimates of interval size by calculating an average between the unimodal signals.

It is possible that the latencies between the modal signals were still within a temporal window of crossmodal integration, whereby modal signals that are temporally asynchronous (up to a certain extent) are still perceived as an integrated stimulus (Spence & Squire, 2003). Summerfield (1987) argues that the time-varying characteristics of the unimodal signals (e.g., noticeable changes in vocal tract in auditory, visual, and vibrotactile information) are one of the primary components under which crossmodal integration occurs. Thus, slowing one of the signals could have been more effective at reducing accuracy than varying the temporal asynchrony between the signals.

Appendix B

In order to assess the effects of stimulus intensity on accuracy, a separate group of 18 individuals (average age = 22 years, 9 male, 4 left-handed) that had an average 10.5 years of musical training (SD = 10.7) participated for course credit. The test stimuli, design, procedure, and statistical analyses were similar to Experiment 1 except that the auditory and vibrotactile signals were not calibrated. It was predicted that calibration would have no significant effect on audio-alone and vibrotactile-alone accuracy scores when compared to the same non-calibrated conditions. Small breaks were given between test conditions.

Results

Accuracy. Accuracy scores were subjected to a 3x2x3x11x2 mixed ANOVA that used condition (audio-alone, visual-alone, vibrotactile-alone), singer (1^{st} or 2^{nd}), repetition (1^{st}, 2^{nd}, 3^{rd}), and interval (1-11 semitone difference) as within-subject variables and order (A-V-T, A-T-V) as the between-subjects variable. No significant effect of condition was found for accuracy scores, $F(2, 30) = 1.92$, $p = .164$, $\eta^2 = .11$, $MSE = 34.85$ (See Figure 11). Singer had a significant effect on accuracy scores, $F(1, 15) = 5.63$, $p = .031$, $\eta^2 = .27$, $MSE = 26.98$. A pairwise comparison revealed that the second singer ($M = 2.38$) received significantly greater accuracy scores than the first singer ($M = 2.58$). Maulchy's test indicated that the assumption of sphericity was violated for interval, $\chi^2(54) = 176.94$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .210$. No significant effect of interval was found, $F(2.095, 31.429) = 2.01$, $p = .149$, $\eta^2 = .12$, $MSE = 133.88$. No effect of order or repetition was found. The covariate of music expertise did not have a significant effect on accuracy scores, $F(1, 15) = 3.80$, $p = .070$, $\eta^2 = .20$, $MSE = 85.51$.

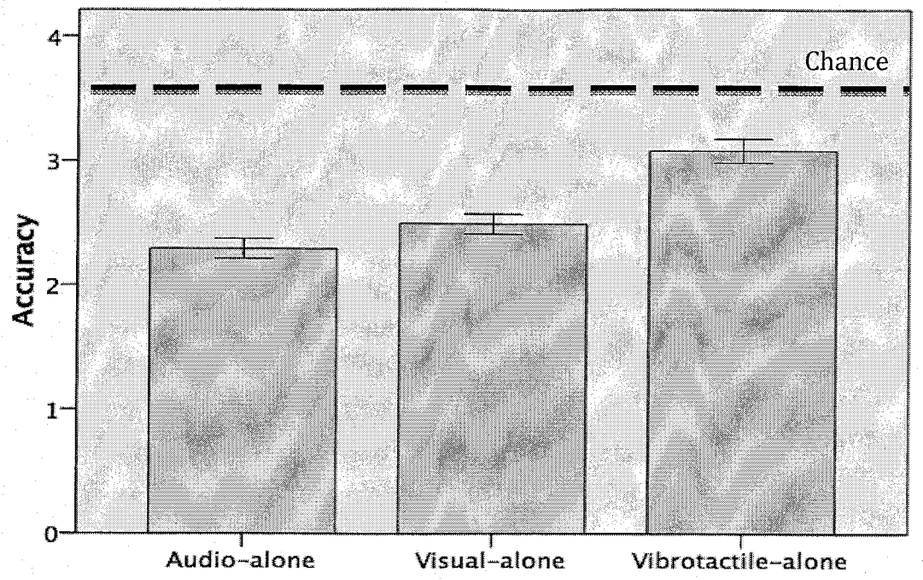

Figure 11. Mean accuracy scores across conditions for non-calibrated trials. Error bars represent +/- 1 standard error.

Precision. Precision scores were subjected to a 3x2x11 mixed ANOVA that used condition (audio-alone, visual-alone, vibrotactile-alone) and interval (1-11 semitone difference) as the within-subjects variable and order (A-V-T, A-T-V) as the between-subjects variable. No significant effect of condition was found for precision scores, $F(2, 30) = 1.63$, $p = .213$, $\eta^2 = .10$, $MSE = 2.55$ (see Figure 12). The covariate of music expertise had a significant effect on precision scores, $F(1, 195) = 15.01$, $p < .001$, $\eta^2 = .07$, $MSE = 15.17$. Planned contrasts revealed that higher levels of music experience significantly improved precision for audio-alone, $t(195) = 3.63$, $p < .001$, and vibrotactile-alone conditions, $t(195) = 2.65$, $p < .001$. Maulchy's test indicated that the assumption of sphericity was violated for interval, $\chi^2(54) = 85.19$, $p < .001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity, $\varepsilon = .787$. A significant effect of interval was found, $F(4.641, 69.620) = 3.26$, $p = .012$, $\eta^2 = .18$, $MSE = 3.77$. A simple plot revealed no

trends in accuracy across interval sizes.

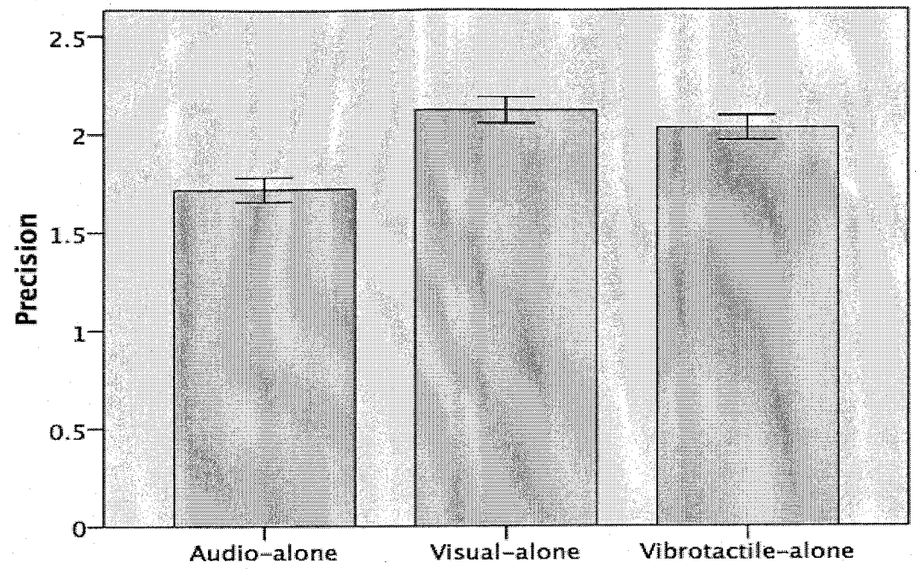

Figure 12. Mean precision scores across conditions for non-calibrated trials. Error bars represent +/- 1 standard error.

Comparison to Experiment 1. The results do not reveal any differences in accuracy based on condition, unlike Experiment 1 where audio-alone received the best accuracy scores, followed by visual-alone, and followed by vibrotactile-alone. Independent t-tests revealed significant differences in accuracy between the experiments for audio-alone, $t(3364) = 2.50$, $p = .012$, and visual-alone, $t(2618.026) = 2.83$, $p = .005$, but not vibrotactile-alone, $t(3364) = 1.896$, $p = .058$. Audio-alone accuracy scores were significantly better when the auditory stimuli were calibrated. Participant groups significantly differed in terms of musical training, $t(20.887) = 2.44$, $p = .023$. Participants had a higher level of musical training in the non-calibration experiment ($M = 10.9$ years) compared to the calibration experiment ($M = 4.1$ years).

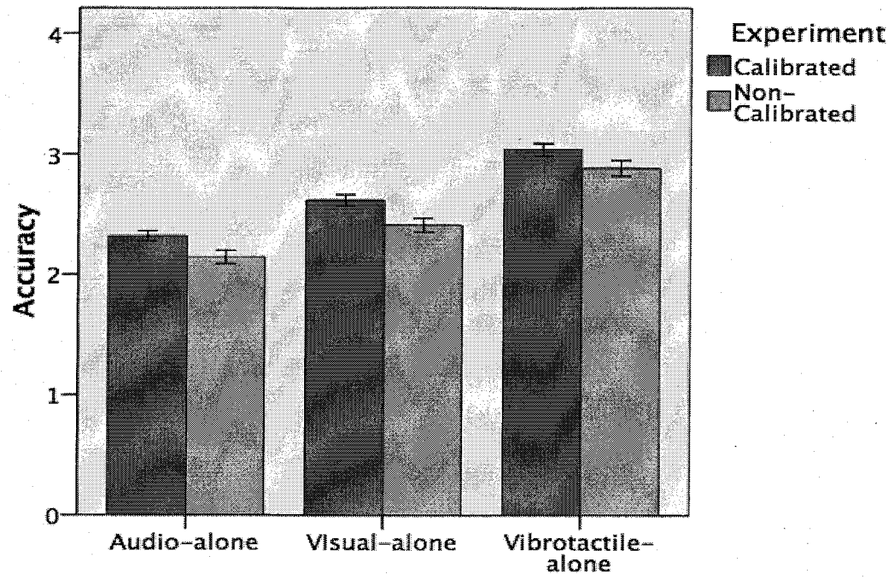

Figure 13. Mean accuracy scores for conditions between calibrated audio-alone and vibrotactile-alone trials from Experiment 1 and non-calibrated audio-alone and vibrotactile alone trials. Error bars represent +/- 1 standard error.

Discussion. The results suggest that auditory estimates of interval size benefit when the auditory stimuli are not intensity-calibrated. Accuracy scores for vibrotactile trials were similar across experiments. However, differences in musical expertise across experiments could explain the significant differences in accuracy for all 3 conditions. These results suggest that pitch discrimination is a primary component of interval size estimation when using auditory and vibrotactile signals.

Appendix C

Chance estimates were calculated as follows. First, all possible responses for a particular interval were subtracted from the interval size. Second, the absolute values were averaged to obtain a chance estimate for that particular interval. An overall chance estimate was calculated by taking an average of all the chance estimates for each interval (see Table 3).

Table 3.
Chance estimates of interval size calculated for each interval.

Interval	**Response**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Chance**
1	0	1	2	3	4	5	6	7	8	9	10	5
2	1	0	1	2	3	4	5	6	7	8	9	4.18
3	2	1	0	1	2	3	4	5	6	7	8	3.54
4	3	2	1	0	1	2	3	4	5	6	7	3.09
5	4	3	2	1	0	1	2	3	4	5	6	2.82
6	5	4	3	2	1	0	1	2	3	4	5	2.73
7	6	5	4	3	2	1	0	1	2	3	4	2.82
8	7	6	5	4	3	2	1	0	1	2	3	3.09
9	8	7	6	5	4	3	2	1	0	1	2	3.54
10	9	8	7	6	5	4	3	2	1	0	1	4.18
11	10	9	8	7	6	5	4	3	2	1	0	5
											Overall Chance Estimate:	3.64

Appendix D

Research on vibrotactile signals suggests that the tactile system is sensitive enough to discriminate across a range of frequencies that capture a significant portion of the frequencies employed in music. Although the diatonic intervals used in each experiment cover the peak frequency of vibrotactile sensitivity (250-300 Hz), it is possible that participants failed to discriminate intervals smaller in range because of limitations due to vibrotactile frequency discrimination. In other words, poor accuracy scores on smaller intervals may have masked potential judgment benefits from the vibrotactile modality.

Holding intensity constant, Pongrac (2008) determined that for anchor frequencies of 250 and 350 Hz, participants required the comparison frequency to be 50 Hz and 80.5 Hz above or below the anchor, respectively. The smallest 4 intervals used for testing in the present study (1, 2, 3, and 4 semitone-range intervals) have frequency differences of less than 80 Hz between the first and second notes, which may have been insufficient to register a just noticeable difference (i.e., the smallest difference in frequency between two notes for which the individual can report the two notes as different). Therefore, these intervals may have all felt the same to the participant.

Using data from Experiment 1, a paired-samples t-test was performed to determine if estimates of interval size in the vibrotactile-alone condition were significantly worse for smaller intervals versus larger intervals. Intervals were split into 2 groups. The first group contained test intervals in which the first and second notes differed by less than 80 Hz (1-4 semitones). The second group contained test intervals in which the first and second notes differed by more than 140 Hz (8-11 semitones). A significant difference was found in accuracy between the interval groups, $t(791) = 6.44$, $p < .001$. However, accuracy scores

for the smaller semitone-range intervals ($M = 2.80$) were significantly better than accuracy scores for the larger semitone-range intervals ($M = 3.66$).

Discussion. This finding suggests that participants were in fact better at judging interval size for smaller intervals versus larger ones. Two explanations can be given for this. The first is that music consists of a higher occurrence of smaller intervals versus larger ones, thereby providing individuals with more extensive experience in discriminating between notes that have smaller differences in pitch. The second explanation is that participants were hesitant to answer with higher estimates of interval size and thus restricted their range to lower estimates, thereby worsening accuracy for larger interval sizes. Based on this result, vibrotactile discrimination for smaller intervals was not considered a concern towards finding any potential benefit of the vibrotactile modality towards musical judgments.

Appendix E

MUSIC QUESTIONNAIRE

Name: _____ Age: _____ Gender: Male / Female
Phone: _____ Email: _____
Are you Right or Left Handed? Right / Left Is English your first language? Yes / No

I. Formal music training:

1. Have you ever taken music lessons (ANY type of lessons count, e.g., high school band class)? Yes / No
   * If **YES**, please continue to #2; If **NO**, please proceed to #4

2. Please indicate your *instrument/voice training*, using a different line for each different instrument or voice:

Instrument/Voice	Individual (years)	Group (years)	RC Grade*	Age at time of lessons
1)				
2)				
3)				
4)				

*If not Royal Conservatory training, what method of training? _____

3. Please indicate your *music theory training* (if any):

Type (e.g., composition)	Individual (years)	Group (years)	RC Grade*	Age at time of lessons
1)				
2)				
3)				

*If not Royal Conservatory training, what method of training? _____

II. Informal music training/current music involvement:

4. Have you ever taught yourself to play an instrument (i.e., without formal lessons on that instrument)?

Instrument	How long played?
1)	
2)	

5. Are you currently active musically (i.e., within the last year)? Yes / No

If 'Yes': ___ Recreational (indicate activity): _____
___ Formal lessons (indicate activity): _____

6. Do you listen to music (circle one)? Yes / No

If 'Yes', how often (e.g., everyday for about 3 hours)? _____
If 'Yes', what type (e.g., classical, rock)? _____

7. What is your favorite type of music? _____
8. Is music important to you? Yes / No If 'Yes', how? _____
9. Do you consider yourself musical? Yes / No / Somewhat
10. Do you have normal hearing? Yes / No

References

Alais, D., Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257-262.

Alcorn, S. (1932). The Tadoma method. *Volta Review, 34*, 195-198.

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

Békésy, G. V. (1949). The structure of the middle ear and the hearing of one's own voice by bone conduction. *Journal of the Acoustical Society of America, 21*(3), 217-232.

Birnbaum, D. M., Wanderley, M. M. (2007). A systematic approach to musical vibrotactile feedback. In *Proceedings of the 2007 International Computer Music Conference*.

Büchel, C., Price, C., Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature, 16*(394), 274-277.

Calvert, G. A., Spence, C., Stein, B. E. (Eds.) (2004). *The Handbook of Multisensory Processes.* Cambridge, MA: MIT Press.

Cook, N. (1998). Analysing Musical Multimedia. Oxford: Clarendon Press.

Davidson, J., Correia, J. S. (2002). Body movement. In *The Science and Psychology of Music Performance*, R. Parncutt and G. E. McPherson (Eds.), 237–253. New York: Oxford University Press.

Dekle, D. J., Fowler, C. A., Funnel, M. G. (1992). Audiovisual integration in perception of real words. *Perception and Psychophysics, 51*(4), 355-362.

Dissanayake, E. (2001). *Homo Aestheticus: Where Art Comes From and Why*. Seattle: University of Washington Press.

Driver, J., Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical transactions of the Royal Society of London: Biological Sciences, 353,* 1319-1331.

Fowler, C. A., Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*(3), 816-828

Fraisse, P. (1974). *Psychologie du rhythme.* Paris: Presses Universitaires de France.

Gescheider, G. A., Bolanowski, S. J., Verillo, R. T. (2004). Some characteristics of tactile channels. *Behavioural Brain Research, 148*, 35-40.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gick, B., Jóhannsdóttir, K. M., Gibraiel, D., Mühlbauer, J. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *Journal of the Acoustical Society of America Express Letters, 123* (4), 72-76. doi: 10.1121/1.2884349

Granström, Björn; House, D.; Karlsson, I. (Eds.) (2002). *Multimodality in Language and Speech Systems.* New York: Springer.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology, 63*(3), 289-293.

Howard, D., Rimell, S., Hunt, A., Kirk, R., Tyrrell, A. (2002). Tactile feedback in the control of a physical modeling music synthesizer. *Proceedings of the 7th International Conference on Music Perception and Cognition*, Sydney.

Howard, I. P., Templeton, W. B. (eds.) (1966). *Human Spatial Orientation*. Wiley: London, England.

Jones, J. A., Jarick, M. (2006). Multisensory integration of speech signals: the relationship between space and time. *Experimental Brain Research, 174*, 588-594.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In J. P. Juslin & J. A. Sloboda (Eds.), *Music and Emotion: Theory and Research*. Oxford University Press, NY: New York.

Karam, M., Nespoli, G., Russo, F. A., Fels, D. I. (2009). Modeling perceptual elements of music in a vibrotactile display for deaf users: A field study. In *Second International Conferences on Advances in Computer-Human Interactions*, 249-254.

Karam, M., Russo, F. A., Branje, C., Price, E., Fels, D. I. (2008). Towards a model human cochlea: Sensory substitution for crossmodal audio-tactile displays. In *Graphics Interface Conference 2008.*

Lerdahl, F., Jackendoff, R. (Eds.) (1983). *A generative theory of tonal music*. Cambridge: MIT Press.

Liberman, A. M. (1984). On finding that speech is special. *American Psychologist, 37*, 148-167.

Liberman, A. M., Cooper, F., Shankweiler, D., Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Liberman, A. M., Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Lovelace, C. T., Stein, B. E., Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research, 17*(2), 447-453.

Macleod, A., Summerfield, Q. (1990). A procedure for measuring audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology, 24*(1), 29-43.

Marshall, M. T., Wanderley, M. M. (2006). Vibrotactile feedback in digital musical instruments. In *Proceedings of the 2006 Conference on New Interfaces for Musical Expression,* 226–229.

Massaro, D. (Ed.) (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, New Jersey: Erlbaum.

Massaro, D. (1989). Review of *Speech perception by ear and eye: A paradigm for psychological inquiry. Behavioral and Brain Sciences, 12*, 741-755.

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748.

Meredith, M. A., Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56*, 640-662.

Mithen, S. J. (Ed.) (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body.* London: Weidenfeld & Nicolson.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Research, 14*(1), 115-128.

Patel, A. D. (Ed.) (2007). *Music, Language, and the Brain.* Oxford: Oxford University Press.

Phillips, D. P., Farmer, M. E. (1990). Acquired word deafness, and the temporal grain of sound representation in the primary cortex. *Behavioural Brain Research, 40*(2), 85-94.

Pongrac, H. (2008). Vibrotactile perception: examining the coding vibrations and the just noticeable difference under various conditions. *Multimedia Systems, 13*, 297-307.

Rosenbaum, L. D., Schmuckler, M. A., Johnson, J. A. (1997). The McGurk effect in infants. *Perception and Psychophysics, 59*(3), 347-357.

Rothenberg, M., Verillo, R. T., Zahorian, S. A., Brachman, M. L., Bolanowski, S. J. (2006). Vibrotactile frequency for encoding a speech parameter. *Journal of Neurophysiology, 95*, 1442-1450.

Russo, F. A., Thompson, W. F. (2005). The subjective size of melodic intervals over a two-octave range. *Psychonomic Bulletin and Review, 12*, 1068-1075.

Saldaña, H. M., Rosenbaum, D. (1993). Visual influences on auditory pluck and bow judgments. *Perception and Psychophysics, 54*, 406-416.

Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304.

Shuster, L. I., Durrant, J. D. (2003). Towards a better understanding of the perception of self-produced speech. *Journal of Communication Disorders, 36*(1), 1-11.

Siegel, J. A., Siegel, W. (1977a). Absolute identification of notes and intervals by musicians. *Perception and Psychophysics*, *21*, 143-152.

Siegel, J. A., Siegel, W. (1977b). Categorical perception of tonal intervals: Musicians can't tell *sharp* from *flat*. *Perception and Psychophysics*, *21*, 399-407.

Sohmer H., Freeman S., Geal-Dor M., Adelman C., Savion, I. (2000) Bone conduction experiments in humans—a fluid pathway from bone to ear. *Hearing Research, 146*, 81–88.

Spence, C., Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 1005-1030.

Spence, C., Squire, S. B. (2003). Multisensory integration: Maintaining the perception of synchrony. *Current Biology, 13,* 519-521.

Stein, B. E., London, N., Wilkinson, L. K., Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience, 8*(6), 497-506.

Stein, B. F., Meredith, M. A. (Eds.) (1993). *The Merging of the Senses.* Cambridge, MA: MIT Press.

Stein, B. F., Meredith, M. A., Wallace, M. T. (1994). Development and neural basis of multisensory integration. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives,* 81-105. Hillsdale, NJ: Erlbaum.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading.* London: Erlbaum Associates.

Tallal, P., Miller, S., Fitch, R. H. (1993). Neurobiological basis of speech: A case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences, 682*(1), 27-47.

Thompson, W. F., Graham, P., Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica, 156*, 177-201.

Thompson, W. F., Russo, F. A. (2004). The attribution of meaning and emotion to song lyrics. *Forum Psychologiczne, 9*, 51-62.

Thompson, W. F., Russo, F. A. (2007). Facing the music. *Psychological Science*, *18*(9), 756-757.

Thompson, W. F., Russo, F. A., Quinto, L. (2006). Preattentive integration of visual and auditory dimensions of music. In *Proceedings of the Second International Conference on Music and Gesture*, RNCM, Manchester, UK.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*, 80-113.

Vroomen, J., de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1583-1590.

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

Wassenhove, V., Grant, K. W., Poeppel, D. (2006). Temporal window of integration in auditory-visual speech perception, *Neurophysiologia*, *45*, 598-607

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19*(5), 515-522.