12-1-2017

MUSIC TO OUR EYES: ASSESSING THE ROLE OF EXPERIENCE FOR MULTISENSORY INTEGRATION IN MUSIC PERCEPTION

Robert Edward Graham
Southern Illinois University Carbondale, regraham85@gmail.com

MUSIC TO OUR EYES: ASSESSING THE ROLE OF EXPERIENCE FOR MULTISENSORY INTEGRATION IN MUSIC PERCEPTION

By

Robert Graham

B.S.O.F., Indiana University, 2007
M.A., Queen's University, 2010
M.A., Southern Illinois University Carbondale, 2014

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in the field of Psychology

Department of Psychology
in the Graduate School
Southern Illinois University Carbondale
December 2017

DISSERTATION APPROVAL

MUSIC TO OUR EYES: ASSESSING THE ROLE OF EXPERIENCE FOR MULTISENSORY INTEGRATION IN MUSIC PERCEPTION

By

Robert Graham

A Dissertation Submitted in Partial

Fulfillment of the Requirement

for the Degree of

Doctor of Philosophy

in the field of Psychology

Approved by:

Dr. Usha Lakshmanan, Chair

Dr. David Gilbert

Dr. Reza Habib

Dr. Eric Lenz

Dr. Jennifer Mishra

Professor Jay Needham

Graduate School
Southern Illinois University
October 30, 2017

AN ABSTRACT OF THE DISSERTATION OF

ROBERT GRAHAM, for Doctorate of Philosophy degree in PSYCHOLOGY, presented on October 30, 2017 at Southern Illinois University Carbondale.

TITLE: MUSIC TO OUR EYES: ASSESSING THE ROLE OF EXPERIENCE FOR MULTISENSORY INTEGRATION IN MUSIC PERCEPTION

MAJOR PROFESSOR: Dr. Usha Lakshmanan

Based on research on the “McGurk Effect” (McGurk & McDonald, 1976) in speech perception, some researchers (e.g. Liberman & Mattingly, 1985) have argued that humans uniquely interpret auditory and visual (motor) speech signals as a single intended audiovisual articulatory gesture, and that such multisensory integration is innate and specific to language. Our goal for the present study was to determine if a McGurk-like Effect holds true for music perception as well, as a domain for which innateness and experience can be disentangled more easily than in language. We sought to investigate the effects of visual musical information on auditory music perception and judgment, the impact of music experience on such audiovisual integration, and the possible role of eye gaze patterns as a potential mediator for music experience and the extent of visual influence on auditory judgments.

108 participants (ages 18-40) completed a questionnaire and melody/rhythm perception tasks to determine music experience and abilities, and then completed speech and musical McGurk tasks. Five auditory stimuli per task were created from spoken and musical (cello and trombone) sounds that ranged incrementally along a continuum from one type to another (e.g. non-vibrato to strong vibrato). In the audiovisual condition, these sounds were paired with videos of the speaker/performer producing one type of sound or another (representing either end of the continuum) such that the audio and video matched or mismatched to varying degrees. Participants indicated, on a 100-point scale, the extent to which the auditory presentation

represents one end of the continuum or the other. Auditory judgments for each sound were then compared based on their visual pairings to determine the impact of visual cues on auditory judgments. Additionally, several types of music experience were evaluated as potential predictors of the degree of influence visual stimuli had on auditory judgments. Finally, eye gaze patterns were measured in a different sample of 15 participants to assess relationships between music experience and eye gaze patterns, and eye gaze patterns and extent of visual on auditory judgments.

Results indicated a reliable “musical McGurk Effect” in the context of cello vibrato sounds, but weaker overall effects for trombone vibrato sounds and cello pluck and bow sounds. Limited evidence was found to suggest that music experience impacts the extent to which individuals are influenced by visual stimuli when making auditory judgments. The support that was obtained, however, indicated the possibility for diminished visual influence on auditory judgments based on variables associated with music “production” experience. Potential relationships between music experience and eye-gaze patterns were identified. Implications for audiovisual integration in the context of speech and music perception are discussed, and future directions advised.

ACKNOWLEDGMENTS

It honestly takes a village.

- Thank you to my Faculty Advisor and Dissertation Committee Chair Dr. Usha Lakshmanan for her invaluable guidance over the years.
- Thank you to the rest of my dissertation committee (Dr. David Gilbert, Dr. Reza Habib, Dr. Eric Lenz, Dr. Jennifer Mishra, Professor Jay Needham) for their feedback and suggestions.
- Thank you to the undergraduate research assistants in Dr. Lakshmanan's Psycholinguistics lab, Rebecca Zilkowski, Tanner Dillon, and Tonya Hamilton for their help running participants and managing data collection.
- Thank you to Dr. Zach Pilot for his help in the lab, and statistical analysis feedback.
- Thank you to my wife Sarah Graham for her unending love and support during this difficult journey.
- Thank you to my parents for their encouragement in following my curiosity and passion.
- Thank you to my participants for their time and effort in completing my study.
- Thank you to the SIU office administrators for their help with participant recruitment.
- Thank you to the Graduate School at Southern Illinois University, Carbondale for supporting my research through the Dissertation Research Assistantship.
- Thank you to the Arthur R. Menendez Memorial Fund for Graduate Student Vision Research for making the eye-tracking data collection possible.
- Thank you to Dr. Jane Smith and the Psychology Department at the University of New Mexico for hosting me during data collection.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

As we navigate our world, we encounter environmental events that produce many different forms of energy and information: the sound of your coworker talking on his phone (too loudly), the smell of freshly baked cookies (if you are lucky), or maybe the feel of dirty water spraying you as a car drives through a puddle nearby (if you are unlucky). We experience and perceive these events through the filter of our sensory receptors, ultimately converting this information into electro-chemical signals processed by our brains. Thankfully, we have gotten pretty good at this as a species, and are able to detect these environmental events through our various senses. Traditionally, this includes sight, sound, touch, taste, and smell, though researchers have proposed more subtle differences in sensory perception (e.g. proprioception, pain, temperature, etc.), making defining what constitutes a “sense” slightly more difficult.

Complicating the sensory conversation further is the fact that frequently, these environmental events produce energy that we transduce via multiple modalities simultaneously. Signals from these modalities do not remain isolated; rather, they interact, coming together to contribute to a deeper and more coherent perceptual interpretation of the event. In this way, we efficiently integrate sensory signals from multiple sensory modes to not only construct more elaborate representations of events, but to also more effectively act upon and make judgments about the world around us. We can combine auditory and visual signals to locate where in space an event occurred. Or, if auditory and visual signals occur too far apart in time (usually exceeding a temporal window of few hundred milliseconds), we may determine that each signal originated from a different environmental event (Navarra et al., 2005). Our prior knowledge of the world also gives us expectations regarding what signals accompany each other, for instance knowing that when your too-loud coworker drops his cell phone on the floor, it should make something like a "thwack" rather than a "splash" sound.

One result of this sort of perceptual expectation we have developed is that information from one sensory modality may influence the way we perceive information from another modality. The possibility of cross-modal influence brings us to the heart of the present study. Here, we attempt to identify domains in which such influence can be observed and measured, and assess the extent to which the integration of multimodal information may depend on the specific domain or be more domain-general within the brain. We also explore the potential for individual experience to shape multimodal perceptual expectations enough to significantly alter the combination and prioritization of information from each modality.

One way to investigate cross-modal influence is to construct a modal mismatch of some sort between types of signals that appear to originate from the same event. The result is that perceivers sometimes experience "illusory percepts" such that they perceive a different signal from what is actually presented. While not limited to audiovisual sensory integration, one of the most well-known examples of this phenomenon has been discovered for speech perception in the context of what is now called "The McGurk Effect."

In an influential study by McGurk and MacDonald (1976), researchers investigated the way in which audio and visual signals are combined in speech perception. The most prominent condition in the study involved participants viewing a film of a woman's talking head repeating (visually) the syllable /ga/, while the audio had been dubbed over with the syllable /ba/. Many participants in this context report hearing the spoken syllable /da/ rather than either of the syllables presented. The implication is that in order to process the information accurately and map it onto familiar speech patterns we expect from those signals, participants “fuse” the signals by combining features of each into a different third signal, /da/, that shares features of both.

The exact mechanisms involved in this particular illusory percept are not entirely understood, but interestingly, this phenomenon often occurs even when participants are aware of the nature of the mismatch in signals, suggesting it is a fairly strong, automatic, and pre-attentive phenomenon (though the limitations of this automaticity will be addressed later). As a result, some researchers have argued that the nature of combining auditory and visual information for speech perception is special, and potentially a result of innate language processing mechanisms. Liberman and Mattingly (1985) have proposed the revised Motor Theory of speech perception, in which they suggest that auditory and visual signals during speech are combined by a perceiver pre-attentively into one coherent “intended phonetic gesture” made up of both signals (auditory and visual articulations), and this is subsequently processed for content. In this way, speech production and perception are proposed to be intimately linked, and different in kind from other forms of auditory or visual perception. They argue that this intimate link is not a learned association, but the product of an innate mechanism that unfolds with development (Liberman & Mattingly, 1985).

Other theories of speech perception differ in varying degrees to the Motor Theory, but most make less bold statements regarding the specialness or innateness of speech perception. The Direct-Realist approach, for example, argues for the importance of the combination of auditory and gestural information for speech processing, but suggest that speech perception is not different from perception in other auditory domains in this way, or even unique to humans (Fowler, 1996). Other researchers (under a broader umbrella of Auditory Theories of speech perception) don’t deny the link between motor production and sound perception, but maintain that acoustic information is sufficient (and that theorized coarticulated motor information is at least not mandatory) for speech sound perception and categorization (Kluender, Diehl, & Killeen, 1987; Diehl, Lotto, and Holt, 2004). The relevance of these theories is addressed in the discussion chapter.

Regarding an evaluation of the Motor Theory of speech perception, the original McGurk effect findings lend support to the idea that auditory and visual signals are combined into a coherent gesture early in the processing timeline, but do little to help examine the other features of the Motor Theory. Even if the McGurk effect were specific to language, does it suggest there are distinct, specialized multisensory integration systems for certain perceptual tasks such as speech perception? Or, do behaviors simply make use of a domain-general audiovisual integration system with varying degrees of efficiency? If language is in fact found to be “special” or privileged for sensory integration, the relative roles of innate and environmental factors need to be explored in other ways, as the traditional speech-based McGurk Effect doesn’t address this claim of the Motor Theory of Speech Perception directly.

Some researchers (Rosenblum, Schmuckler, & Johnson, 1997; Burnham & Dodd, 2004) have found that visual perception influences auditory perception in pre-lingual infants in McGurk-like tasks involving habituation to various audiovisual matched and mismatched stimuli. However, there is still a possibility of experience influencing perception with these infants, especially given the possibility of early linguistic critical periods. Additionally, there is evidence that young children (3-5 and 7-8 years old) demonstrate less overall susceptibility to visual influence on auditory perception, suggesting that it increases with age and use (McGurk & MacDonald, 1976). Ultimately, it remains somewhat difficult to separate experiential from innate abilities in a speech context since language is something we begin experiencing even before birth.

To bypass this conflation of innate and experiential factors in audiovisual integration, one can examine additional perceptual contexts and behaviors in which this type of sensory integration is an important part of perception and production. One such context is that of music, since there exist individuals with a wide range of musical experience from those with none to those with almost life-long experience both producing and perceiving musical sounds. Though it could be sufficient to only be a long-time “receptive” experiencer of music (e.g. watching and listening to performances only), it is likely that actively producing and performing music would instill stronger audiovisual perceptual expectations and cross-modal influence, especially considering the possibility of “mirror neurons” associated with producing and perceiving certain motor behaviors (Molnar-Szakacs & Overy, 2006).

Since there is a wider range of musical experience and ability compared to language experience across humans, investigating audiovisual integration in a musical context provides an opportunity to answer some more subtle questions about the role of experience in multisensory integration. If audiovisual integration is at least somewhat experience-driven, then it may be possible to assess the type and quantity of experience necessary for the sort of cross-modal influence taking place in the McGurk Effect to occur (if an appropriate comparison can be established in music). It may be that an adequate duration (e.g. x number of years) of active experience is necessary to elicit an influence of visual on auditory perception, or just that a certain number of practice/performance hours must be achieved. It is also unclear how generalizable such experience-based sensory integration effects might be. If someone has extensive experience playing the violin, will audiovisual integration processes for violin behaviors and sounds also generalize to other stringed instruments? What about brass instruments?

The goal of the present study was to broadly investigate the role of experience in cross-modal influence for speech and music behaviors, and to attempt answers for some of the subtler questions regarding the extent and type of music experience necessary to facilitate a significant influence of visual stimuli on auditory perception. Further, eye gaze patterns were tracked as a measurable external behavior that may provide insight into the relationship between experience and multisensory integration. We use these approaches to assess the proposed “specialness” of language in the context of multisensory integration, and examine how diverse types of experience might play a role in how we combine auditory and visual information in musical contexts. As a result, this study has a broader impact on the scientific community regarding discussions on multisensory integration, modularity in the brain, learned and innate processes, and identifying observable behaviors that correspond to these processes.

CHAPTER 2

LITERATURE REVIEW

To provide sufficient context for the present study, we review a broad range of related research in this section. We start by exploring the criteria used to determine if sensory information from different modalities should be interpreted as originating from the same source. From here, a discussion of the lower level (neural and cognitive) sensory integration processes is provided. We then offer an overview of variations on the McGurk effect since the present study aims to demonstrate one such variation, followed by highlights from relevant eye-tracking research that helped to inform analytical decisions. Multisensory integration is then discussed as it relates to music, leading into an examination of the study that motivated the research questions and methodology of the present study. Finally, hypotheses for the present study are provided.

Binding Bimodal Events with Time and Semantics

Temporal synchrony and integration. One necessity for integrating information from two different sensory modalities (e.g. auditory and visual) is that they occur relatively close together in time. A "temporal window" can act as a sort of filter when determining whether signals of two different sensory modalities originated from the same environmental event. However, interpreting this window is made somewhat more difficult by the underlying nature of auditory and visual signals and processing. As Recanzone (2009) points out, auditory signals (sound) from environmental events take longer to travel to human sensory receptors than visual signals (light), while the physiological processing time tends to be faster in the auditory system than the visual system, so there are some inherent disparities in place already. Interestingly, judgments of perceived synchrony between auditory and visual signals depend on which modality is presented first. When the visual signal arrives slightly before the auditory, participants are more likely to judge the two signals as synchronous than if the auditory signal arrives slightly before visual (Recanzone, 2009).

In cognitively “normal” humans, if auditory and visual signals arrive at the sensory receptors within about 200-300 ms of each other, they are more likely to be perceived as originating from the same event, and thus integrated into a coherent audiovisual signal (though the likelihood still diminishes within this window the farther apart the two signals are). While these temporal synchrony windows are likely mostly universal and innate, they can be altered slightly through experience, resulting in slightly different tolerances for what might be detected as synchronous or asynchronous. Navarra et al. (2005) demonstrated some flexibility within this 300 ms window by asking adults to monitor asynchronous audiovisual signals, including both speech and musical stimuli. They found that with habituation, the adults would “recalibrate” audiovisual signals within the temporal window of about 300 ms such that they were less sensitive to changes in synchrony within the window. However, for asynchronies beyond (e.g. 1000 ms), no such recalibration occurred.

Lewkowicz (2010) expanded on this finding by investigating temporal synchrony and sensory integration in infants. The researchers found that short-term experience in the form of habituation with infants was enough to alter responses to audiovisual stimuli of varying degrees of synchrony. Specifically, when infants were habituated to synchronous audiovisual speech signals (syllables), they subsequently only detected asynchronies of greater differences, approximately 666 ms apart. Interestingly, when habituated to the asynchronous signals 666 ms apart, they appeared to be more sensitive to differences in synchrony, detecting both the difference between 666 ms and 0 ms (synchronized), as well as the difference between 666 ms and 366 ms. Similar findings were also found for non-speech signals, suggesting a more domain-general mechanism (Lewkowicz, 2010). This increase in sensitivity (or reduction of the synchrony window) from exposure to asynchrony in infants is somewhat different than what was reported for adults, as in the Navarra study. Lewkowicz posits that this is a result of a much longer history with synchronous audiovisual events for adults, resulting in a perceptual bias toward unified audiovisual events. In sum, it appears that temporal synchrony is important for multisensory integration, but that it may depend on both short-term and long-term (developmental) experience.

Semantic matching and integration. Through extensive histories of interacting with the environment, humans have acquired knowledge and expectations about how sounds and corresponding objects/organisms are matched in the real world. One result of this experience is that people tend to have enhanced (e.g. faster) responses for objects/organisms presented with both auditory and visual information available. Suied, Bonneel, and Viaud-Delmon (2009) found that participants responded significantly faster to bimodal presentations of visual stimuli (in this case, a telephone and frog) paired with their corresponding sounds compared to unimodal visual or auditory presentations alone – an effect that has come to be known as the "redundant signal effect." Of note is that this effect was only found if the auditory and visual signals were congruent, or semantically matched. Interestingly, for the incongruent pairings, an interference effect (e.g. an increase in reaction time compared to unimodal presentations of stimuli) was only found when the target was visual and the distractor was auditory, but not when the target was auditory with a visual distractor. This suggests a possible asymmetry in the filtering of irrelevant or distracting sensory information between auditory and visual signals (Suied et al., 2009).

To demonstrate the importance of experience to these audiovisual semantic judgments, other studies have looked to object familiarity to indicate the necessity of prior knowledge. One such study used functional Magnetic Resonance Imaging (fMRI) to measure differences in brain activity for the presentation of unfamiliar objects/sounds, familiar (e.g. animal images/sounds) congruent objects/sounds, and familiar incongruent objects/sounds (Hein et al., 2007). It was found that integration of unfamiliar object images and sounds (artificial and arbitrary images/sounds) involved the inferior frontal cortex, which they interpret as reflecting the learning of new audiovisual associations. They also found activity in the inferior frontal cortex for the integration of familiar but incongruent images/sounds, with additional activity in the posterior superior temporal sulcus. Finally, for familiar and semantically congruent audiovisual stimuli pairings, activity was again found in the posterior superior temporal sulcus, but additional activity was present in the superior temporal gyrus (Hein et al., 2007). These imaging findings present an interesting spectrum of overlapping activity across familiar/unfamiliar and congruent/incongruent bimodal stimuli.

A study utilizing event-related potentials (ERP) found support for distinct types of neural activity based on semantic matching. Liu, Wang, and Li (2011) looked at ERP responses to semantically matched, moderately matched, or mismatched audiovisual signals from different types of environmental events – a wine glass falling and shattering, a person exiting a room and closing a door, a fireworks display, etc. Compared to the semantically matched condition, they found stronger N400 responses (negativity approximately 400 ms post-event) in both the moderately matched and mismatched conditions. They interpreted this activity as potentially reflecting a connection process (or failure to do so) between the perceived actions/sounds and semantic memory (Liu et al., 2011). They also found a P600 component (positivity 600 ms post-event) in the semantically moderately matched compared to matched conditions, which they have interpreted as a potential evaluation or reanalysis process regarding the incoming information (Liu et al., 2011).

Taken together, these results suggest different (and potentially enhanced) human responses and activity to semantically matching (congruent) bimodal signals compared to mismatched and unimodal signals, suggesting that semantic matching can act as a filter for multisensory processing in the same way that temporal synchrony does.

The "How," "When," and "Where" of Multisensory Integration

Neural activity and multisensory integration. To discuss the "how" of multisensory integration, it is important to explore the nature of multisensory convergence at the neuronal level. Multisensory integration can be thought of as perhaps the overall process of combining information from multiple modalities into a coherent signal, utilizing both top-down and bottom-up processes (see a proposed sequence of events in Figure 1).

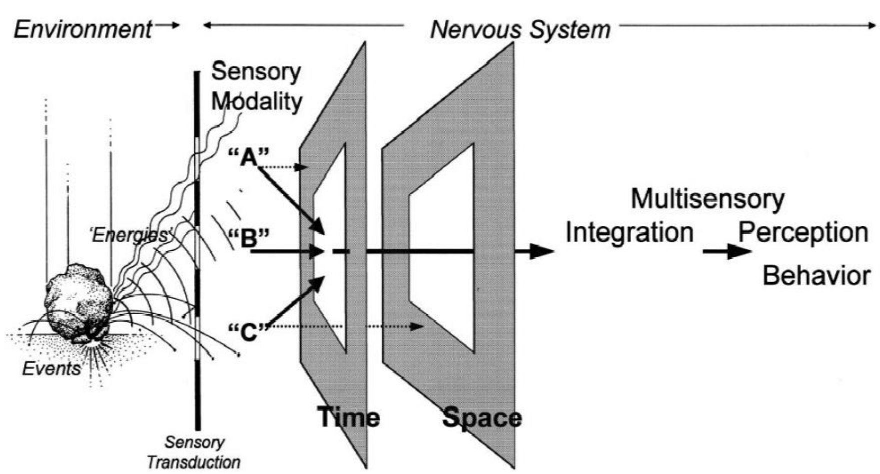


Figure 1. The sequence of multisensory processing, including a spatial-temporal window/filter facilitating integration (Meredith, 2002, p. 35).

On the other hand, multisensory convergence is more of a bottom-up process occurring as either “neuronal convergence” or “areal convergence” (Meredith, 2002). With neuronal convergence, signals from an environmental even travel from distinct sensory receptor organs and converge on shared neurons for multiple sensory systems to influence one another. Conversely, areal convergence involves information from multiple modalities converging on similar brain areas, though not necessarily sharing individual neurons (Figure 2).

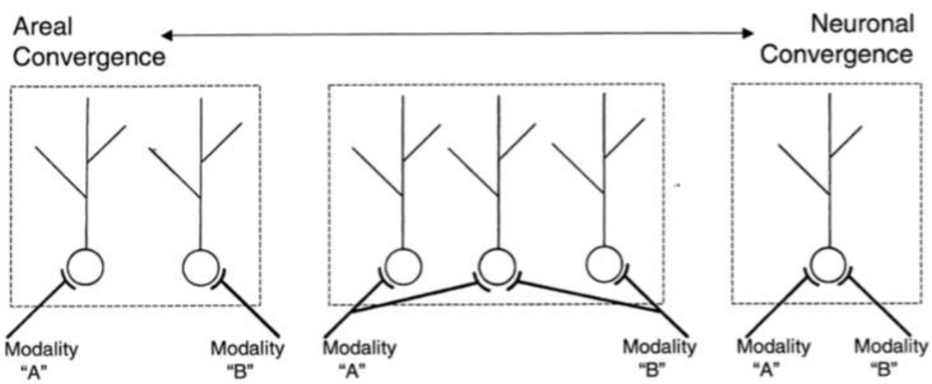


Figure 2. Areal and Neuronal convergence of signals from two different modalities (Meredith, 2002, p. 33).

Of importance, particularly with neuronal convergence, is that responses to one mode of sensory information can now be modified by activity elicited by another mode, allowing modal interactions to take place at the neuron. As a result, responses to information from multiple senses at individual bimodal or even trimodal neurons may be very different in degree and nature from unimodal neural responses. Signals can be enhanced or inhibited with more than one sensory modality input compared to separate single-mode inputs. At any individual multisensory neuron, multiple excitatory and inhibitory signals are converging and combining to determine an overall net excitatory or inhibitory response. In this way, for example, you could have compounding excitatory auditory and visual signals converging on a bimodal neuron, resulting in an enhanced response greater than the sum of those two signals alone – or, "super-additive" enhanced activity (Kayser, Lakatos, & Meredith, 2012). However, you can also have excitatory-inhibitory convergence, which may be important when information from one modality may need to be enhanced while information from another modality may need to be inhibited, as with selective attention (Meredith, 2002).

Importance for the neural activity is often placed on the type of net response dictated by these converging signals, but with excitatory-inhibitory convergence, there are ultimately many signals that remain "subthreshold," without being expressed as part of the net response for a particular modality. Kayser et al. (2012) argue that multisensory influences really exist along a continuum, from unimodal neurons that only respond to one type of sensory information, to "classical" bimodal neurons that display "suprathreshold" responses to sensory information from multiple modalities (even presented separately). Between these points on the continuum, you can then have those that behave like some neurons found in the auditory cortex of macaque monkeys, which respond to sounds (presented alone) but not visual stimuli, though still show an enhanced (or reduced) response for the simultaneous presentation of stimuli from both modalities (Kayser et al., 2012).

Another proposed component of multisensory integration at the neural level has less to do with enhancing or suppressing signal convergence, and more to do with neural activity in the form of consistent or sometimes synchronized oscillatory patterns at various frequency bands (Senkowski, Schneider, Foxe, & Engel, 2008). Senkowski et al. (2008) suggest that beta band oscillations are enhanced when sensory stimuli onsets occur close together in time, such that beta oscillations, temporal contiguity (of stimuli presentation), and multisensory salience are all somewhat linked. Von Stein, Rappelsberger, Sarnthein, & Petsche (1999) further found synchrony in the form of oscillatory coherence (a constant relationship between oscillatory signals from two different brain regions) in the beta band between electrodes at temporal and parietal sites in the context of semantic integration for auditory and visual object processing. In regard to consistent oscillatory activity, Senkowski, Molholm, Gomez-Ramirez, and Foxe (2006) found enhanced evoked beta oscillations for audiovisual stimuli (compared to auditory or visual alone) that allowed for the prediction of shorter reaction times to the stimuli. Relevant to the present study, Kaiser, Hertich, Ackermann, Mathiak, and Lutzenberger (2005) found enhanced gamma band activity when illusory auditory percepts were induced in the McGurk Effect (See the McGurk Effect and Beyond section for a discussion on this phenomenon) similar to activity found in an auditory mismatch task. This indicates that the gamma band activity may represent perceptual changes facilitated by crossmodal interactions. Though just a representative selection of findings is provided here, there is a wide body of research that supports the importance of multisensory neural signal convergence and neural oscillation for a reasonable (though still incomplete) model of multisensory integration.

Attention and audiovisual multisensory integration. The answer to “when” in the processing stream multisensory integration takes place is somewhat unclear, since studies have demonstrated conflicting findings regarding how (or if) multisensory integration and attention interact. This has resulted in three different explanations: an “early integration” framework, a “late integration” framework, and a “parallel integration” framework (Koelewijn, Bronkhorst, & Theeuwes, 2010). The early integration framework proposes that multisensory integration is separate from attention, in that multisensory integration can occur at an early sensory level preattentively. However, there may be cases in which integration influences attention, for example with bimodal cues perhaps priming attention at a higher level. The late integration framework suggests, instead, that unimodal attention can impact individual sensory inputs, allowing for integration at a later stage into one percept or event. Under this model, multisensory integration relies on attention to enhance signals such that they may be integrated later. One possible way to reconcile seemingly conflicting findings supporting these frameworks is to suggest that multisensory integration takes place at multiple processing stages, as the parallel integration suggests.

Calvert and Thesen (2004) propose that multisensory integration may occur at early or late stages depending on the resources available and the requirements of the task. One possibility under this framework is that attentional resources may facilitate the integration of near-threshold multisensory neural events, such that integration only occurs at stages sensitive to top-down influences. Conversely, supra-threshold events may not require attention for integration, such that they may integrate automatically at early stages (though attention could still impact later integration as well) (Koelewijn et al., 2010). In this way, it is possible to have integration occurring at multiple stages of processing in parallel. This possibility is also supported by ERP data acquired by Molholm et al. (2002). The researchers compared ERP responses to auditory, visual, and audiovisual stimuli, and examined the difference between audiovisual and auditory plus visual as an indicator of the occurrence of multisensory integration (AV – (A+V) = Integration). They found this integration ERP signal at various time points, ranging from about 46 ms after stimulus onset (which was simultaneous with visual cortical processing activity, often the earliest activity detected) to about 200 ms post stimulus onset (Molholm et al., 2002).

Localization of audiovisual multisensory integration. If the parallel multi-stage processing model of integration is accurate, then it also complicates the discussion on “where” in the brain audiovisual sensory integration takes place. In the study by Molholm and colleagues (2002), they found the early audiovisual integration activity corresponding with electrodes over the dorsal parieto-occipital scalp area, which they interpret as activity at an early stage of the dorsal visual stream. However, at slightly later audiovisual interaction responses, the activity appears to progress elsewhere: at about 120 ms, the signal becomes more right superior temporal and left centro-parietal, and at about 180 ms the audiovisual interaction is over the left central scalp, thought to reflect sensory motor integration (Molholm et al., 2002).

Another difficulty in identifying exactly where audiovisual integration takes place lies in the variety of ways neurons process sensory information from different modes. Some researchers have found sites in the posterolateral lateral suprasylvian cortex of cats (an area typically involved with visual motion processing) that have both bimodal and unimodal neurons with “multisensory properties” (Allman & Meredith, 2007). They identified one “bimodal zone” in which both visual and auditory stimuli were independently effective in activating neurons. Meanwhile, they identified a neighboring site possessing neurons that selectively activated to visual stimuli (and not auditory), but demonstrated enhanced responses when the visual signal was simultaneously accompanied by an auditory signal (thus categorized as a subthreshold multisensory region). Clemo, Sharma, Allman, and Meredith (2008) expanded on this finding by determining where exactly the auditory signals communicating with these regions were coming from. They ultimately identified several auditory cortical areas projecting signals to these regions: the primary auditory cortex, the secondary auditory cortex, the dorsal zone, and the field anterior ectosylvian sulcus (Clemo et al., 2008). These findings indicate that various auditory cortical regions communicate with and potentially influence visual processing, particularly involving visual motion processing.

Regarding the beta band oscillations discussed in the previous section, some studies have suggested the importance of beta band oscillations for audiovisual integration in general, and the McGurk Effect more specifically. Keil, Müller, Ihssen, and Weisz (2011) found a connection between beta band activity in the left superior temporal gyrus (lSTG) and the presence of a "fusion" effect (or illusory percept) in the McGurk task. In particular, they determined that the perception of a fusion effect is preceded by increased beta band activity in the lSTG (and a few other regions), and also found increased right frontal beta activity, decreased coupling of the lSTG with right temporal areas, and increased coupling of the lSTG with frontoparietal areas (Keil et al., 2011). In this way, localizing activity in this context is less about absolute activity increases in specific regions, and perhaps more about changes in the relationships between various regions.

Some studies have also presented findings supporting the role of the superior colliculus (SC) in audiovisual integration. Activity in the SC is often associated with eye movement and head orientation toward objects in space, but research on multisensory integration has indicated that the deeper layers of the SC contain many neurons that respond to information from different sensory modalities (Stein, 2012). A study by Maravita, Bolognini, Bricolo, Marzi, and Savazzi (2007) investigated activity in the SC in relation to the "redundant signal effect" discussed earlier – specifically, reaction times to simultaneous audiovisual signals are typically faster than to auditory or visual signals alone. They examined SC activity using different colored stimuli, since signals resulting from red stimuli typically project to the SC, while signals from purple/blue stimuli do not. Results favored the involvement of the SC, as a redundant signal effect was found with the presentation of red stimuli, but not with blue/purple stimuli (Maravita et al., 2007).

An additional study investigated speech and non-speech signals in the SC, using synchronous and asynchronous presentations of bimodal (audiovisual) stimuli contrasted with auditory and visual signals presented alone. Researchers used fMRI to measure activity across the brain, identifying regions that displayed superadditive (AV > A+V) response enhancement (for synchronous presentation) and depression (for asynchronous presentation) (Calvert, Hansen, Iverson, & Brammer, 2001). They found that both synchronous and asynchronous presentations of bimodal non-speech stimuli facilitated enhanced responses compared to auditory or visual stimuli presented alone in several different brain regions, similar to speech stimuli. In particular, the most profound differences in response were found in the SC, though other regions demonstrated differences as well, including cortex within the superior temporal sulcus, intraparietal sulcus, insula, and superior and ventromedial frontal gyri (Calvert et al., 2001).

Finally, research with interesting implications to the present study involves audiovisual integration activity in the superior temporal sulcus (STS). Stevenson and James (2009) used fMRI to compare brain activity to multisensory speech and tool stimuli. While they identified different regions if interest in the STS for speech and tool stimuli, they found similar patterns of multisensory enhancement across speech and tool stimulus responses compared to distinct auditory and visual regions of interest. The researchers uncovered similar patterns of “inverse effectiveness,” arguing that multisensory enhancement increases as stimulus saliency decreases. In this case, stimulus saliency is measured by using different signal-to-noise levels, such that there was less superadditive (excitatory multisensory) activity when stimuli were presented in low noise conditions, but more superadditive activity when stimuli were presented in high noise conditions. This suggests a stronger reliance on multisensory cues when environmental events are more ambiguous. More importantly for the present study, these findings support a model of distinct regions for speech vs. tool audiovisual integration, but similar processes for each, suggesting that audiovisual integration may be similar across behaviors without necessarily being innate.

McGurk Effect and Beyond

Development and audiovisual illusions. As discussed previously, the McGurk Effect suggests that sensory information from one mode (e.g. vision) can influence perception in another mode (e.g. audition), such that a fusion of signals may occur, resulting in an illusory percept. One way to examine the role of experience in multisensory integration (and susceptibility to illusory percepts) is to conduct studies with infants involving audiovisual perception and illusions. While infants old enough to respond in some meaningful way (through eye fixations or head turns) have acquired some limited experience of the world, they provide the closest thing to a population without "experience," so it becomes more plausible that a process or mechanism is innate if it is found in infants.

Rosenblum et al. (1997) examined the presence of a McGurk Effect in 5-month-old infants by habituating them to audiovisual presentations of the syllable /va/. They then expose them to audio /ba/ - visual /va/ stimuli, and found that they appear to generalize their habituation to this condition, but not to an audio /da/ - visual /va/. These results suggest infants of this age are visually influenced in the same way adults are on such a task. The researchers find that it more strongly supports the possibility of innate audiovisual integration mechanisms, but that experience cannot be ruled out. A similar study by Burnham & Dodd (2004) either habituated

4.5-month-old infants to audio /ba/ - visual /ga/ (experimental group) or to matching audiovisual /ba/ (control group). They found that in subsequent auditory-only trials, the experimental group treated /da/ and /ða/ (tha) as familiar (without having heard them previously) and /ba/ as novel. The control group showed no such familiarity/generalization to any of the novel stimuli. Again, these findings largely support the presence of a McGurk Effect in prelinguistic infants.

ERP data gathered by Kushnerenko, Teinonen, Volein, and Csibra (2008) also suggest visual influence on auditory perception in 5-month-old infants. They measured brain activity to audio /ba/ - visual /ga/ and audio /ga/ - visual /ba/ compared to the congruent versions of those syllables. Adults typically will not demonstrate a mismatch response to audio /ga/ - visual /ba/ (suggesting successful "illusory" integration), but will to audio /ba/ - visual /ga/. ERP data suggested that this was also the case with infants, such that only the audio /ba/ - visual /ga/ resulted in additional activation approximately 290 ms following sound onset across frontal and temporal areas, suggesting this incongruent pairing was processes as a mismatch. The researchers acknowledge, however, that such results could be due to the salience of the visual stimuli, and that such findings may not be universal across languages since some languages (e.g. Japanese) may provide less distinctive visual information (Kushnerenko et al., 2008).

Finally, Tremblay et al. (2007) examined the McGurk Effect across a broader age range, using participants ranging from 5- to 19-years-old. They investigated the presence of the audiovisual illusion with both speech and non-speech stimuli. The non-speech stimuli included the "Shams illusion," in which a single visual flash can be perceived as two flashes if accompanied appropriately by two successive sounds, and the "fusion effect," in which two separate visual flashes can be fused into one when paired with a single auditory signal. The most interesting result from this study involved divergent findings for the speech and non-speech tasks regarding development. For the speech task, there appears to be a weaker McGurk Effect (less likely for cross-modal influence to occur as it does in adults) at younger ages (participants from 5 to 9 years old) than for the older participants. However, for the two non-speech tasks, the illusory effects were present and consistent across all age groups. While the nature of the speech and non-speech tasks are necessarily a bit different in regard to how participants are asked to respond, this finding suggests a possibility that either experience or biological maturation may play a larger role in audiovisual speech integration than in non-speech audiovisual integration.

One recent finding that suggests a role for experience comes from Proverbio, Massetti, and Zani (2016) comparing (instrumental) musicians and non-musicians on a speech McGurk task spanning several different Italian phoneme presentations. More specifically, they found no significant McGurk Effect overall for musicians on incongruent audiovisual stimuli compared to auditory presentations of the same phonemes, while the non-musician group did demonstrate a McGurk Effect for tasks including alveolar-nasal (/na/), velar-occlusive (/ka/ and /ga/), and bilabial (/pa/ and /ba/) phonemes. They attribute this reduced McGurk Effect in musicians to possible neurological (e.g. greater connectivity between cortical areas) changes stemming from musical training that influence broader audiovisual integration. This directly challenges the suggestion that audiovisual integration for speech is unique and innate (Liberman & Mattingly, 1985).

Linguistic and attentional variations on the McGurk Effect. In the time since McGurk and MacDonald (1976) discovered the so-called McGurk Effect, many variations have been conducted to assess the generalizability of such a finding, as well as limitations and influencing factors. One such idea has surfaced in the form of the “native-foreign language effect,” which suggests that speakers may experience a stronger McGurk effect in a foreign language compared to their native language, due to a greater reliance on visual information (lip movement) in assessing auditory content. Hayashi and Sekiyama (1998) explored this effect in native speakers of Chinese and Japanese, though it should be noted that there was a wide range of Japanese proficiency in the Chinese participants, but the Japanese participants did not speak Chinese.

The researchers had native speakers of both Chinese and Japanese pronounce various syllables (e.g. /ba/, /pa/, /na/, and /ga/) in their respective languages for use in the tasks, and stimuli were constructed to include congruent and incongruent audiovisual stimuli. All participants completed audio-only, video-only, and audiovisual tasks (resented in both Japanese and Chinese) in which they reported the syllables spoken. As suggested by previous studies (Sekiyama, 1997), both Chinese and Japanese participants showed weaker McGurk Effects in their native languages than Americans tend to demonstrate with English (indicating the presence of language-specific influences on the McGurk Effect). The Japanese participants did appear to show stronger McGurk Effects for the Chinese stimuli, which is consistent with the 'native-foreign language effect," but the Chinese participants did not show a difference in effect between Chinese and Japanese stimuli. A couple explanations for the difference between language groups include the difference in foreign language proficiency between the groups, as well as the fact that Chinese is a tone language, which may foster a stronger reliance on auditory cues (Sekiyama, 1997; Hayashi & Sekiyama, 1998). Regardless of the reason, these findings at least suggest that the McGurk effect may vary in prevalence depending on the type of language spoken by an individual, as well as the extent of bilingual proficiency.

Other variations on the McGurk Effect have sought to assess the automaticity of the effect. One such study by Soto-Faraco, Navarra, and Alsius (2004) did this by utilizing a speeded classification paradigm in which reaction times to the first syllable of word-like stimuli are slowed when the second (distractor) syllable varies inconsistently. The stimuli included different version of the spoken non-words “tabi,” “tobi,” “tadi,” and “todi” in an audiovisual context. The researchers confirmed that in a matching (congruent) audiovisual task, the second syllable interfered with responses to the first. In two subsequent versions of the task, the goal was to facilitate (and then eliminate) a McGurk-like auditory illusion such that despite a mismatch in audiovisual signals (e.g. audio “tobi” paired with visual “togi,”) the “perceived” (but illusory) second distractor syllable can still interfere with response to the target syllable. The results confirmed this effect, offering evidence of interference of an auditory percept on target syllable perception. This finding supports the notion that the McGurk Effect is fairly automatic, given that the illusory syllable impacted their ability to process the target syllable, despite a lack of awareness of the illusion.

Other studies have suggested a stronger role of attention in multisensory integration, which seems to indicate a less automatic and more conscious multisensory integration process. Research by Alsius, Navarra, Campbell, and Soto-Faraco (2005) indicates that under conditions of high attentional load, one’s multisensory integration processes may function slightly differently. They investigated this possibility by having participants complete the classic McGurk task in the context of a dual-task paradigm. In one experiment, they completed the McGurk task concurrent with a visual repetition task (determining if objects presented visually repeated) and in another experiment, participants completed a concurrent auditory repetition task (determining of sounds presented repeated). Other participants were exposed to the concurrent tasks, but not required to respond regarding them. They found that in both experiments, participants concurrently performing both tasks had significantly reduced visual signal influence on auditory perception, so reduced McGurk Effect. The necessity of attention for the McGurk effect suggests that this audiovisual binding may not be as automatic or pre-attentive as once believed, however the researchers propose that attentional demands impact the unimodal sensory processing, prior to binding (Alsius et al., 2005). It is also possible that, as discussed earlier, integration is taking place in parallel at early and late stages of sensory processing, allowing for interactions at various stages.

To address whether attention is acting at a unimodal (pre-binding) or bimodal (at binding) level, Alsius, Navarra, and Soto-Faraco (2007) again looked at the impact of high attentional demands on audiovisual integration, but did so this time by adding concurrent tactile stimulation (through tapping devices). Again, they found that when attentional demands are increased, even in different sensory modes, that visual signals have less impact on the processing of auditory signals in a McGurk task. This suggests that the attentional processes involved in impacting the McGurk Effect are likely at a more general sensory integration level rather than a modality-specific level, since the extra attentional demand is not on a sensory mode already in use for the McGurk task. This possibility is in line with the possible of parallel integration across early and late stages of processing (see Calvert & Thesen, 2004).

Eye-tracking, Expertise, and Attention

Researchers have been interested in what eye movements can suggest about cognitive processes for a long time, but the technology and methodology utilized to measure such movements has advanced rapidly in the last couple of decades. The present study involves equipment that uses a video-oculography method of capture, because of its minimally invasive interface. This approach uses remote (desktop) cameras that capture pupil features and a corneal reflection by bouncing near infra-red light off the eyeball, and then using the relationship between reflections to estimate gaze location in space (Janthanasub & Meesad, 2015). Researchers can then track measurements like the location and duration of gaze "fixations," which are periods of apparent synchronized eye immobility between larger "saccades," or rapid gaze shifts. Interestingly, these fixations actually involve very small eye movements (historically categorized as microsaccades, ocular drift, and tremors) across a point of focus (Rucci, McGraw, & Krauzlis, 2016). For the purposes of the present study however, we will treat fixations as mostly stable points of focus within a scene for ease of analysis.

Gaze fixations allow researchers to determine direction of visual attention, though this can in turn be used to infer other cognitive processes and perceptual strategies. Reingold, Charness, Pomplun, and Stampe (2001) investigated a relationship between eye movements and expertise in the context of chess. Given a check-detection task (in which participants scanned a miniature two-dimensional board to assess the presence of a "check"), greater expertise was found to be related to fewer fixations compared to less skilled players, but a greater number of fixations between chess piece images compared to less skilled players. Interestingly, when the chess piece images were replaced with letters representing each piece (e.g. Q for queen), this difference between experts and non-experts diminished, suggesting domain-specific enhanced processing for chess experts rather than broader perceptual advantages. The researchers argue that the presence of more fixations between pieces supports a hypothesis by Chase and Simon (1973) suggesting that expertise in this context is partially associated with more internal representations of patterns and relationships among the pieces as "chunks." This is potentially applicable to other domains as well, and indicates that visual attention for experts may not always be directed to the most "relevant" piece of visual information if there is a broader context to consider.

A meta-analysis of research investigating eye-tracking and expertise in various domains (Gegenfurtner, Lehtinen, & Säljö, 2011) found comparable results, with experts displaying fewer total fixations than intermediate and novice participants across domains, with more fixations on task-relevant information and fewer fixations on task-redundant information. They characterize these findings as supporting an information-reduction hypothesis put forth by Haider and Frensch (1999), which differs slightly from the Chase and Simon chunking hypothesis in that it emphasizes efficient attention to relevant information paired with the ability to inhibit or avoid redundant information.

Gurler, Doyle, Walker, Magnotti, and Beauchamp (2015) used eye-tracking technology to investigate the relationship between eye gaze and susceptibility to the McGurk Effect. In general, they found that individuals experiencing the effect were more likely to fixate on a speaker's mouth during the task, but note that these individuals were still more likely to experience the effect even on trials in which they did not fixate on the mouth, suggesting it is not as direct a relationship as one might guess. They propose that individuals experiencing the McGurk Effect more frequently may have a history of weighting visual information more strongly than others, so that even if they are only viewing the lip movements in their slight periphery, this still may influence them more. In this way, the fixations on lip movements would be more a signal of their cognitive processing history than a direct cause of the effect. Additionally, the researchers acknowledge other studies in which the McGurk Effect is lessened when participants' attention is directed toward distractors, and note that eye-gaze is ultimately an imperfect proxy for visual attention, so there will surely be variability in such relationships.

When exposed to more real-world perceptual speech difficulties (e.g. not intentionally deceptive stimuli such as the McGurk Effect), perceivers tend to rely on visual lip movements more strongly when speech is either somewhat ambiguous or the source of speech is unclear to aid in comprehension. Yi, Wong, and Eizenman (2013) found that under clear perceptual contexts (single speaker, low background noise), participants were able to report utterances with comparable accuracy despite variable fixation locations up to about 10° away from the center of the mouth. However, when an additional speaker is added and/or background noise is increased, participants tended to fixate their gaze near the mouth more when able to shift freely, and accuracy of responses dropped significantly when participants were required to have a fixed gaze location beyond 2.5° away from the mouth. This suggests that in day-to-day life, perceivers naturally alter their gaze strategy to utilize more lip and mouth visual information under suboptimal perceptual conditions. This is important for the present study, which utilizes somewhat ambiguous auditory stimuli combined with more distinct visual stimuli.

Multisensory Integration in Music

Visual performance, emotion, and music perception. With advances in recording technology, humans today likely experience music in a purely auditory context most of the time. However, visual performances of music contribute to the emotional content of the music, and this visual component is lost with audio recordings. Molnar-Szakacs and Overy (2006) suggest that as with the proposed motor theory of speech perception, music involves a close coupling of perception and production regarding structured musical information. In particular, the mirror neuron system – which is proposed to demonstrate neural activity both when an organism is perceiving an action being performed, as well as when an organism is performing that action – is proposed as a candidate for facilitating such connections between music production and perception. They take this a step further, however, and propose a common neural substrate for music, language, and motor functions, as studies of language disorders have shown interactions between these processes. For example, Melodic Intonation Therapy is a music therapy technique used in speech recovery, which is highly imitative, and incorporates rhythmic motor movements as well (Sparks, Helm, & Albert, 1974). Additionally, Molnar-Szakacs and Overy suggest that music and language can both be considered temporally-unfolding systematic hierarchical structures of communication in some respect (Lerdahl and Jackendoff, 1983; Patel, 2003). They propose that humans comprehend all communicative signals (regardless of sensory mode or type of activity) in terms of the motor actions behind the signal, and in regard to the *intention* behind the action. I believe that some researchers would be hesitant to couple a processing of motor actions with a processing of intention, but Molnar-Szakacs and Overy argue that musical action in the form of motion (both physical motion and the movement of musical pitches) also conveys emotion to a certain extent (Molnar-Szakacs & Overy, 2006). Thus, the way in which we attribute emotion to a musical performance is coupled to our emotional experience of that music.

Research by Petrini, McAleer, and Pollick (2010) suggests that for assessments of affect in music, auditory signals may influence visual affect judgment more than visual signals influence auditory affect judgments. Participants were exposed to musical excerpts of a drummer or saxophonist playing in an audio-only, visual-only, or audiovisual (matched and mismatched) conditions, and were asked to judge the perceived emotion and rate the strength of the emotion. They found that auditory signals had a greater impact on interpretation of visual affect, though this was primarily in the saxophone condition. In a subsequent study, they found that having emotionally incongruent (mismatched) audio and visual signals when judging the visual affect worsens performance on the task, though only if the audio and video signals originate from the same instrument. This suggests that there may be a categorical or semantic filter requirement that must be met first to indicate the signals are originating from the same source before the signals can interact to influence affect judgments.

Thompson, Graham, and Russo (2005) investigated emotional expression in music by conducting case studies on the use of facial expression, body movement, and gesture in filmed performances by B. B. King and Judy Garland. In the performance by B. B. King, they found that his facial expressions functioned both to indicate his own emotional states, and also to emphasize the character of the music (e.g. dissonance and "blue" notes). Additionally, his facial expressions functioned much more as an extension of his own guitar sounds than as an extension of the music as a whole (including accompanying ensemble instruments). They also inferred certain things from his mouth movements, for instance that a "wince" indicates he is doing technical work. Judy Garland, on the other hand, included gestures to accompany the lyrics of the song, but also includes body movements to highlight certain aspects of the music. When she reached a tonal modulation in the song, she "boldly" walked forward to emphasize the significance.

Thompson et al. (2005) subsequently examined the influence of visual performance cues on auditory perception of music regarding both musical content and affect. In one such study, they had participants rate clips of the B. B. King performance based on auditory musical dissonance, which they described to participants as "occurring when the music sounded discordant (i.e. conflicted or negative) and in need of some sort of resolution" (Thompson et al., 2005). Some participants made the judgments from audiovisual stimuli, while other participants had audio-only stimuli, with stimuli chosen to be either neutral or dissonant in character. In general, participants exposed to audiovisual stimuli had greater differences in dissonance judgments between the dissonant and neutral stimuli compared to the audio-only group, suggesting the visual component of the performance contributed to the overall “sense” of dissonance. Additionally, as an example of visual influence on affect judgment, the researchers found that facial expressions influenced affect judgments during major vs. minor note interval production by a singer. Taken together, their series of experiments demonstrated that the visual signals of a performance (e.g. facial expressions, gestures, and bodily movements) influence perceivers’ music experience at both perceptual and emotional levels.

Beyond dissonance and affect judgments, other researchers have found that visual signals in music can influence perceived musical note duration. Researchers had participants watch videos of percussionists playing long or short notes, paired with long or short auditory notes, to determine how the visual signal of the note being played on the instrument might influence how long the participants judge the note to be. They found that perceived auditory note duration varied more strongly based on the visual signal presented, suggesting that longer musical visual gestures influence notes to sound longer (an illusory percept) due to the multisensory integration taking place (Schutz & Lipscomb, 2007).

Multisensory integration and musical expertise. Recent studies have found that music experience influences the way in which someone perceives and integrates audiovisual information in a musical context, however the nature of this influence is not always clear. Paraskevopoulos, Kraneberg, Herholz, Bamidis, and Pantev (2015) used magnetoencephalographic records to assess brain activity and connectivity between musicians and non-musicians during audiovisual musical tasks. Participants were asked to assess congruency between pitch height, and a disk presented at various heights on a screen. Behaviorally, musicians were able to discriminate between congruent and incongruent trials with greater accuracy than non-musicians. Additionally, brain activity for musicians suggested enhanced processing, with greater connectivity between cortical areas than non-musicians during the task, and with more activity around the right temporal cortex for multisensory integration, as well as around the left inferior frontal cortex for detecting "abstract" audiovisual incongruency (Paraskevopoulos et al., 2015). The pattern of activity also suggests to the researchers that the non-musicians relied more strongly on visual cues than musicians, however it is unclear if this is generalizable to other musical contexts given the abstract nature of the stimuli. In a related study with similar stimuli, Paraskevopoulos, Kuckenbuch, Herholz, and Pantev (2012) found that musicians had a greater sensitivity to incongruent audiovisual stimuli in this sort of task than non-musicians, as indicated by plasticity in superior frontal gyrus, visual cortex, and right secondary auditory cortex.

Musacchia, Sams, Skoe, and Kraus (2007) also found some differences in the ways musician vs. non-musician brains respond to music and speech stimuli (e.g. the spoken syllable /da/ for speech, and a synthesized bowed cello note for music). While they found that both groups generally had enhanced brainstem responses for audiovisual relative to unimodal auditory or visual stimuli, musicians had earlier and larger responses (interpreted as more "robust" pitch encoding) than non-musicians for both speech and music stimuli, and saw greater change in activity between unimodal and audiovisual stimuli. Interestingly, the driving factor behind this enhancement is not always clear, as enhancement was more consistently found to be related less to music "ability" and more to current frequency and consistency of music practice and performance (Musacchia et al., 2007).

While Musacchia and colleagues found that music experience (i.e. training) was associated with pitch processing, a study by Tiippana, Viitanen, and Kivimäki (2013) supposedly found an effect of "musical aptitude" (as determined by modules from multiple musical assessment batteries) on only non-speech audiovisual integration. They tested 10-year-old children on a speech-based McGurk task, and a non-speech audiovisual task previously discussed called the Shams illusion, in which flashes of light perceived are influenced by auditory beeps presented. They found no substantial difference in how lip movements impacted auditory perception across participants (regardless of music experience/aptitude) in the McGurk task, but did find that children with high musical aptitude were less susceptible to the Shams illusion, suggesting that they had a shorter temporal synchrony window for multisensory integration in the context of non-speech stimuli. While Tiippana et al. (2013) refer to musical aptitude as the main difference between groups here, they clarify that after controlling for musical training (greater or less than 1 year of music lessons/training), acknowledge that musical training likely plays a large role in facilitating a weaker Shams illusion, and is likely a large contributing component of what they call musical aptitude (Tiippana et al., 2013). Regardless of the musical training/aptitude distinction, the results suggest that as experience selectively impacts the two audiovisual tasks utilized, a fully domain-general audiovisual integration system appears unlikely.

An interesting study by Hasegawa et al. (2004) explored the role of experience in the context of observed piano playing movements. They used fMRI to measure brain activity in groups of non-pianists, less-trained pianists, and well-trained pianists when watching a pianist's hands (with no audio). The hands would either slide across the keys with no tapping movements (control period), or press keys reflecting familiar pieces, unfamiliar pieces, or random sequences (task periods). Behaviorally, well-trained pianists were able to identify the familiar pieces while the other two groups were not. However, most interestingly, the well-trained pianists also showed increased activity in the left planum temporale during the familiar, unfamiliar, and random sequence task periods. This is especially fascinating because this area of the brain is considered part of the auditory association cortex, involved with audiovisual integration, and particularly lip reading in speech. This part of the brain also appears to activate from reading written language and reading music, suggesting a sort of mapping of visual information to corresponding complex auditory patterns (Hasegawa et al., 2004). Ultimately, this suggests (importantly for the present study) that individuals with extensive enough experience in a specific behavior are equipped to map visual-motor signals involved in the behavior to auditory processes (given that it occurs regardless of familiarity), indicating a greater likelihood of cross-modal influence as experience increases.

Lastly, a study by Proverbio, Calbi, Manfredi, and Zani (2014) used ERP to study audio-visuomotor processing based on instrument-specific expertise. They had groups of violinists and clarinetists watch videos of a violinist and clarinetist playing the same piece of music with similar pitch, intensity, and rhythm. In half of the stimuli, auditory and visual signals were incongruent regarding pitch (e.g. the visual "note" played and the auditory "note" played do not match.) They found that experts watching their own instruments (e.g. a violinist watching the violinist perform) not only process visual information regarding their own instrument faster than others (in the form of an N170 ERP component), but that only experts viewing their own instruments elicited an N400 ERP component in regard to incongruent audiovisual information (suggesting a sort of violation detection). The researchers suggest the presence of "audiomotor mirror neurons" that may encode both musical gestures and sounds (Proverbio et al., 2014). This has important implications for the present study, such that in the context of watching musicians perform, other musicians of the same instrument may process visual signals (articulatory gestures) differently than others, emphasizing the importance of instrument-specific experience.

Motivating study

A study by Saldaña and Rosenblum (1993) expanded on the linguistic McGurk Effect to determine if a visual influence on auditory perceptual judgments might be identified in music that mirrors the speech effect. In the primary experiment, they attempted this by having participants complete auditory-only and audiovisual perceptual judgment tasks. In the auditory-only task, participants listened to a musician pluck a cello string (pulling the string with a finger) or bow a cello string (drawing a horse hair bow across a string to vibrate it), along with three artificial sounds created to bridge the spectrum from a pluck sound to a bow sound. Participants were asked to rate (on a sliding scale) the extent to which each auditory stimulus sounded like a "plucked" or "bowed" sound. In the audiovisual task, the participants heard the same auditory stimuli, but this time paired with video stimuli of a cellist either plucking a string or bowing a string. Participants provided a discrepancy rating between auditory and visual stimuli (with 0 in the middle being congruent, up to a 5 both to the left and right, indicating which type of sound they heard and degree of discrepancy), and again rated the extent to which each auditory stimulus sounded like a "plucked" or "bowed" sound. Participants were aware of the dubbing procedure, and thus the potential for mismatching audiovisual signals. The goal was to determine if pairing the auditory stimuli with different videos influenced the perceptual categorical judgments of the varying auditory stimuli. While the visual stimuli were found to influence auditory judgments slightly, the effect was much smaller than that found in an analogous speech-based task.

While the results seem to suggest weaker McGurk Effect-like multisensory interaction (and thus allow for the possibility that speech audiovisual integration is different in kind from audiovisual integration in other contexts), the present study sought to improve and expand upon theirs in several ways. In addition to having a small sample size of participants, Saldaña and Rosenblum (1993) did not account for musical experience. This has the potential to be an important factor, as it is likely that familiarity with the instrument, or musical exposure more generally, could impact how someone perceives and combines the auditory and visual stimuli. Thus, the present study assessed musical experience across all participants, and made an effort to recruit participants ranging from little musical exposure to extensive musical experience across a variety of instrument types. It was not predicted that music experience would impact performance on a speech-based McGurk task, as evidence regarding music experience on speech perception is somewhat mixed here (See earlier discussions of Musacchia et al., 2007, and Tiippana et al., 2013).

Two different instrument conditions (cello and trombone) were included, made up of multiple tasks utilizing the different articulatory techniques and timbral qualities of each instrument. This allows an examination of how audiovisual integration processes resulting from music experience (if found) generalize across instruments and techniques. Additionally, eye-tracking technology was utilized (though in a more exploratory manner) to determine if there are any relationships between auditory perceptual judgments, musical experience, and eye gaze patterns.

Main Study Hypotheses

1. *Null Hypothesis*: Across the main participant sample, participants' auditory stimulus ratings will not vary significantly based on video stimulus pairing for speech, cello, and trombone tasks.

   Alternative Hypothesis: Across the main participant sample, participants' auditory stimulus ratings for more ambiguous sounds will differ significantly depending on video stimulus pairing for speech, cello, and trombone tasks.

Prior evidence discussed in the literature review suggests that visual indicators of sound production influence the perception of sounds perceived as belonging to the same event (through temporal and semantic synchrony). This has been well documented in speech with variations on the McGurk effect, and there is evidence for this in music both in the context of timbre/note articulation (Saldaña & Rosenblum, 1993) and note duration (Schutz & Lipscomb, 2007).

2. *Null Hypothesis*: Music "production" experience will not significantly predict the extent to which someone's auditory judgment is influenced by visual stimuli in a speech-based McGurk task.

   Alternative Hypothesis: Music "production" experience will significantly predict the extent to which someone's auditory judgment is influenced by visual stimuli in a speech-based McGurk task.

While recent findings (Proverbio et al., 2016) suggest musicians are less susceptible to visual influence in a speech McGurk task than non-musicians, it was not anticipated that this would be the case here (and thus the null hypothesis would not be rejected), partly based on other past findings (Tiippana et al., 2013), and partly due to the difference in experimental design, since a wide range of musical experience is represented in this study.

3. *Null Hypothesis*: The extent to which visual stimuli influence auditory perceptual judgments in the cello and trombone tasks will not be related to musical experience.
   Alternative Hypothesis: Music "production" experience and instrument-specific experience will significantly predict the extent to which visual stimuli influence auditory perceptual judgments on the cello and trombone tasks.

It is predicted that more moderate familiarity with a particular instrument is associated with an influence of visual cues because these individuals would know the basic mechanics of the instrument, but that more expert musicians are less likely to be influenced due to extensive auditory experience with the instrument sounds.

Eye-Tracking Hypotheses

4. *Null Hypothesis*: Musical experience will not predict eye gaze patterns (including gaze location and fixations per second) on the cello vibrato task.
   Alternative Hypothesis: Participants in the eye-tracking sample with string instrument experience will look significantly longer at the cellist's left hand (source of vibrato) and have significantly fewer fixations per second than individuals without string instrument experience.

The sample size for this portion of the study is small, so a formal evaluation of hypotheses pertaining to eye-tracking data is not possible. However, it is expected that those with experience playing a string instrument will have prior knowledge of the mechanics of vibrato, and thus know where to look for the source of this sound feature when considering the extent to which the video and audio match. Further, the research presented earlier on expertise and eye movements (Reingold et al., 2001) suggest that experts, while spending more time looking at relevant visual information, will also display fewer fixations than non-experts.

5. *Null Hypothesis*: Gaze location will not significantly predict the extent to which participants are influenced by visual stimuli when making auditory judgments in the cello vibrato task.

   Alternative Hypothesis: Time spent looking at the cellist's left hand (source of vibrato) will significantly predict the extent to which participants in the eye-tracking sample are influenced by visual stimuli when making auditory judgments in the cello task.

Again, the eye-tracking sample size for this study is small, so clear evaluation of this hypothesis is difficult. Regardless, past findings still inform predictions for the current study. Based on recent eye-tracking research with the traditional speech McGurk effect (Gurler et al., 2015), it is predicted that individuals who spend more time looking at the visual "source" of the sound are more likely to be influenced by visual indicators of sound production in making auditory perceptual judgments.

CHAPTER 3

METHOD

Experimental Design Overview

This study sought to assess how the integration of auditory and visual information may differ based on experience, particularly in the context of music. Music was used because of the broad range of experience individuals possess when compared to a behavior like language. As such, multiple measures of "music expertise" were gathered to determine what might best predict someone's audiovisual integration judgments. Participants representing a wide range of musical experience were recruited, a music background questionnaire was administered, and a computer-based music perception measure was utilized to offer an objective measure in addition to the self-report questionnaire.

Beyond assessing musical expertise, participants completed computer-based tasks to determine the extent to which judgments about auditory stimuli are influenced by accompanying visual stimuli. A speech condition was included to determine if participants differed in how their judgment of speech sounds was influenced by watching lip movements. Though some new research has found differences in susceptibility to the McGurk Effect based on musical experience (Proverbio et al., 2016), we did not predict to find those differences for our speech task. Two music conditions were implemented to assess if visual information influences auditory categorization in a musical context, and if this influence varies based on musical experience. However, it is unclear what type of musical experience might be necessary to make an individual more or less susceptible to such influence, so multiple tasks were used on two different musical instrument conditions. In this way, we can establish how generalizable musical experience might be for different types of audiovisual integration. For example, a cellist watching a cellist perform might integrate the audiovisual information from the performance differently than another type of musician or non-musician. However, it might be that just being a “strings” player (the category of instrument in which a cellist fits) might result in more similar audiovisual integration to a cellist than a “winds” or “brass” instrument player. Thus, “cello” and “trombone” conditions were used to examine a double dissociation for specific types of music experience, such that it would be less about being a musician overall, and more about being a certain type of musician for these tasks.

Finally, eye-tracking data were collected from a sample of 15 individuals on the cello vibrato task, in addition to the questionnaire and PROMS (Profile of Music Perception Skills) task information they completed. The purpose here was to investigate a relationship between gaze patterns and susceptibility to visual influence on auditory judgment, and potentially a relationship between gaze patterns and different types of musical experience. While not allowing for an entirely causal conclusion, it offers an attempt at assessing why differences between participants may exist, if found.

Participants

Upon notice of approval by the Human Subjects Committee at Southern Illinois University, Carbondale (SIU), 108 participants (based on a statistical convention by Green, 1991) ranging from 18-40 (*M* = 22.29, *SD* =4.99) years old were recruited largely from the campuses of SIU, and the University of New Mexico (UNM) in Albuquerque, New Mexico. Students enrolled in introductory psychology classes at SIU and UNM had the opportunity to participate for research points (a course requirement), and students enrolled in other music and psychology courses were offered extra credit for participation in the study. Participants were also recruited through campus and community music organizations via email.

Materials

Questionnaire and PROMS. Participants completed a background questionnaire and modules of the Profile of Music Perception Skills (Law & Zentner, 2012), or PROMS, through a dedicated URL hosted by the University of Innsbruck. Questionnaire items were implemented at the beginning of the profile, followed by the music perception modules. The questionnaire included questions on general/demographic data (e.g. age, handedness, language use, hearing/vision correction, attention/learning disabilities), as well as questions on music experience. Information regarding music listening/viewing experience (e.g. how much individuals listen to music, attend concerts, watch digital music performances) and music performance experience (e.g. instruments played, starting age, duration, frequency, intensity, etc.) was gathered (see Appendix A; note that the questions were presented in a different format electronically). The subsequent music perception modules included melody and tuning modules, in which participants were asked to determine if a) a melody presented in midi tones is the same or different across presentations, and b) if a musical chord presented in midi tones is in- or out-of-tune to the same degree across presentations. Questionnaire and PROMS data are stored on the University of Innsbruck servers but are owned by the PI of this study, and are accessible to download at any time as Excel or SPSS files.

Experimental tasks. The audio and video music stimuli for the experimental tasks were recorded in a studio room in the Department of Radio, Television, and Digital Media at SIU Carbondale, while the speech stimuli were recorded in a room in Santa Fe, New Mexico.

For the musical audiovisual tasks, the video framed the torso of the cellist and trombonist such that hands and upper body were visible, while faces were out of frame. For the speech recordings, only the lower half of the face was used in an attempt for similarity across music and speech tasks. These recordings were incorporated into E-Prime 2.0 (Psychology Software Tools, 2012) and OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) tasks programmed specifically for this research by the PI. Participants viewed the visual stimuli on one of several computer screens, while listening to the auditory stimuli through Sony headphones provided by the experimenter.

Speech stimuli. The speech stimuli involved a male speaker vocalizing consonant-vowel pairs (/va/ and /ba/), which were subsequently edited to create three composite auditory stimuli of the contrasting sounds, such that an incremental continuum of five sounds was established ranging from /va/ to /ba/. This involved shortening the onset of a /va/ sound to be progressively closer to a /ba/ sound.

Cello stimuli. Regarding the stimuli for the cello condition, musical instrument sound samples (from an online sound library) were initially considered as auditory stimuli due to the greater degree of control they would offer for modification and alteration, but a more natural sound, along with a greater ability to pair with the visual stimuli, were prioritized. In one task, the stimuli included recordings of a cello string being plucked (when a cellist uses his or her finger to pull and release the string – this is called pizzicato in classical music), or a cello string being bowed (when a cellist draws the hair of a bow across the string to vibrate it). For this task, hybrid sounds were constructed to span the continuum from “pluck” to “bow.” This involved a combination of altering sound onset and decay, as well as layering of different sound types. In another task, notes on the cello were played with or without a “vibrato” technique, in which the pitch of a musical note purposefully fluctuates around the intended note, usually cycling at a consistent rate. This can occur at faster/slower speeds (changing the duration of cycles) or narrower/wider “amplitudes” (changing the extent to which the pitch varies in frequency from the intended note). The varying vibrato stimuli were naturally produced (e.g. not layered, etc.) with the intention of spanning the continuum from a non-vibrato sound to a very strong vibrato sound.

Trombone stimuli. For ease of comparison, the auditory trombone stimuli mirrored those used in the cello vibrato task in that they were recorded, naturally, to span the continuum from non-vibrato to strong vibrato. Visually, however, these techniques are achieved somewhat differently on trombone than on cello. To have a clearer visual indicator of articulation, the trombone vibrato was created by modulating the right slide hand of the trombonist. This method is employed more selectively, often for “jazzier” music, while more “traditional” vibrato involves more jaw and mouth movements. Since the cello condition is focused on the hand movements of the musician, this task was constructed to also focus on hand movements.

In pilot testing, participants (on average) rated the constructed auditory stimuli for all tasks as spanning the sound continuum as intended (suggesting good validity), though not always in equidistant “steps,” or similarly across tasks. For example, auditory-only vibrato ratings for the cello task increased almost linearly from non-vibrato to strong vibrato, while ratings for the cello pluck/bow task increased drastically from pluck to the middle sounds, then increased at a much slower rate to the full bow sound.

Eye-tracking. A smaller sample of 15 participants completed the cello vibrato task on an ASUS K501UW-AB78 15.6-inch laptop connected to a GazePoint GP3 eye-tracking camera. A 4^{th} generation iPad was used as a remote monitoring screen for the experimenter by connecting it to the computer and using the Duet Display app, which allows the use of iPads and iPhones as screen extensions for computer displays. As the camera takes up space on the keyboard of the laptop, participants entered responses via a Logitech Bluetooth keyboard. Figure 3 depicts the eye-tracking layout (without some of the cords). The PI of this study trained on and wrote an eye-tracking equipment protocol for an Applied Science Laboratories EYE-TRAC6 D6 camera, but due to technical difficulties, data collection was never carried out on this system. Participants in the eye-tracking sample completed the task at a quiet meeting table in the Santa Fe Public Library, while those not included in the eye-tracking sample completed the experiment on computers in lab spaces on the campuses of SIU and UNM.

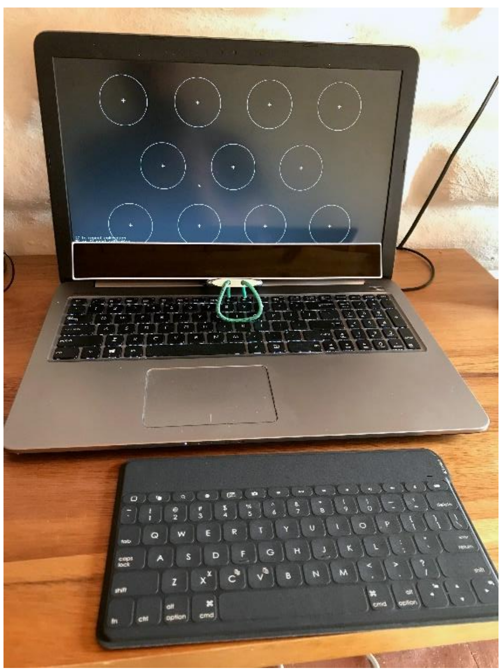


Figure 3. Computer with eye-tracking camera and remote keyboard following calibration.

Procedure

A summary of experiment components and tasks is provided in Table 1. Counterbalancing of components and tasks was used to reduce potential fatigue, practice, and order effects.

Table 1

Experiment Conditions and Order

Component/Condition	Task/Section	
Questionnaire	Demographic Information Music Experience	
PROMS	Melody Tuning	
Speech Condition*	*Auditory-only Task***	*Audiovisual Task***
Cello Condition*	*Auditory-only Tasks*** Pizzicato Task Vibrato Task	*Audiovisual Tasks*** Pizzicato Task Vibrato Task
Trombone Condition*	*Auditory-only Tasks*** Vibrato Task	*Audiovisual Tasks*** Vibrato Task

*Note***.** * Indicates counterbalancing between components. ** Indicates counterbalancing within component.

Questionnaire. Participants began by completing the background questionnaire, which included questions regarding general demographic information, as well as a detailed section on music listening and performance experience (Appendix A).

Profile of music perception skills (PROMS). To supplement the questionnaire, all participants completed a portion of the Profile of Music Perception Skills (PROMS) online as a more easily quantified measure of music perception ability and experience. Specifically, they completed the melody and tuning modules, as these were recommended by the PROMS researcher as sensitive but different indicators of music perception skills.

Speech condition. Participants completed an auditory-only task and an audiovisual task as part of the speech condition. The auditory-only task was comprised of 15 trials involving the auditory presentation of the spoken syllables /ba/ and /va/, including edited versions of these syllables to create a five-point continuum of stimuli ranging from /ba/ to /va/. As such, participants heard each of the five stimulus types three times over the course of the section in randomized order. Participants completed five practice trials with each of the five stimulus types to gain familiarity and were informed about the task procedure. For each trial, they listened for the presentation of one of the syllable stimuli. They were subsequently asked to rate how much the presented syllable sounds like /ba/ or /va/ by selecting a number on a scale ranging from 1 (sounded most like /ba/) to 100 (sounded most like /va/), before continuing on to the next trial. Due to programming constraints, this differed (for all tasks) from the method used in the motivating study (Saldaña & Rosenblum, 1993), which involved a sliding scale.

In the audiovisual task, participants were presented with the same set of auditory syllable stimuli ranging from /ba/ to /va/, but each spoken syllable was paired with a video of a human face that either articulated the syllable /ba/ or the syllable /va/ through mouth and lip movement. Each of the five spoken syllables was paired with the /ba/ video three times and with the /va/ video three times, for a total of 30 trials. Participants were instructed to watch and listen to the stimulus pair for each trial, and first provide a “discrepancy” rating assessing how well the audio and visual components matched to establish that they were paying attention to both audio and visual components of the stimuli. They did this by selecting a number on a scale from 1-5, with 1 indicating the A/V components don’t match at all, and 5 indicating that they fully match. Next, they were asked to rate how much the audio syllable sounded like /ba/ or /va/ by selecting a number on a scale ranging from 1 (sounded most like /ba/) to 100 (sounded most like /va/), based *only* on the auditory syllable presented, regardless of the visual articulation.

Cello condition. Participants completed auditory-only and audiovisual tasks emphasizing a variety of articulatory and timbral features of cello sound. In the auditory-only pluck/bow task, participants were informed about the task and presented with examples of the auditory stimuli. This task included the pluck/bow stimuli described in the materials section, and 3 artificial composite audio stimuli created by editing pluck and bow sounds together to span a continuous range from one sound (pluck) to the other (bow).

As with the speech condition, each (of 5) stimulus was presented 3 times, for a total of 15 experimental stimuli. In the audiovisual task, these auditory stimuli were paired with video of a cellist either plucking or bowing the string, for a total of 30 stimuli (15 auditory stimuli with one video and 15 with the other). The participants were again instructed to provide a discrepancy rating for the extent to which the audio and video matched or mismatched (on a scale from 1-5), and rate on a scale from 1-100 the extent to which the auditory stimulus sounded most like a plucked sound (1) or most like a bowed sound (100).

Within the cello condition, there was an additional task making use of the vibrato technique described in the materials section, utilizing audio and video stimuli of a note played either with wide vibrato or without any vibrato, as well as three audio stimuli spanning the range between these extremes. In all other procedural details, this task was identical to the pluck/bow cello task.

Trombone condition. For comparison purposes, the trombone tasks also used the vibrato technique described in the cello condition section, though visually, this technique is typically achieved somewhat differently on trombone than on cello. Since the cello condition is focused on the hand movements of the musician, the vibrato task was constructed to also focus on hand movements, and thus the “slide vibrato” technique discussed in the materials section was used.

The trombone condition (for both the vibrato and glissando tasks) was similar in design to the cello and speech conditions, with 5 auditory stimuli per task made to range from one sound (e.g. full vibrato) to another (no vibrato). In the auditory-only task, participants had 15 trials in which they heard the stimuli and were asked to rate them on a scale from 1-100, with each end of the scale representing fully one sound or the other. In the audiovisual task, the same auditory stimuli were paired with two videos depicting a trombonist performing one sound or the other, for a total of 30 stimuli. Participants were asked to provide a discrepancy rating for each audio-video pair (on a scale from 1-5), and also to rate the auditory stimulus on the same scale used in the auditory-only task.

Eye-tracking. Eye-tracking data were collected for 15 participants on the cello vibrato audiovisual task to determine if there is a relationship between gaze patterns and musical experience, or gaze patterns and susceptibility to visual influence on auditory judgment. Before starting the eye-tracking task, each participant was asked to adjust their seating and head location so that they would be comfortable with minimal movement during the task. The GP3 camera was then positioned to align with participants’ eyes, and participants completed a 9-point calibration task to establish best point-of-gaze estimation for the camera. Participants were then asked to look at various crosshairs on the screen to verify that the calibration worked properly. If necessary, adjustments were made to lighting and positioning, and calibration was completed again. If calibration was acceptable, then the participant was asked to start the cello vibrato task.

Post-experiment. Following the completion of the research study, participants were debriefed regarding the purpose of the study (verbally and with a written document), and invited to ask any questions they had regarding the purpose and nature of the study. While no formal manipulation check was conducted, many participants reported that they gained some insight into what the study was assessing (influence of visual stimuli on auditory perception) during the study. However, as many people susceptible to the linguistic McGurk Effect still experience the effect even when told about it in advance, this was not thought to be a problem for comparison across language and music tasks.

Data Analysis

Participant decisions and exclusions. This study was primarily concerned with the varieties of experience that might impact audiovisual integration in a musical context. Given the wide range of possible music experience, and the knowledge that multiple predictors may have varying relationships with the dependent variables, a mix of principal component analyses, multiple regressions, t tests, and MANOVAs were utilized. Based on criteria ($N > 50 + 8p$, where p = the number of intended variables) proposed by Green (1991) regarding sample size for regression, a sample of 108 individuals was obtained for the main analyses (not including eye-tracking data).

Exclusion of participant data was determined by a few different criteria. Participants were asked about hearing loss that might inhibit their ability to complete the task, but no participants reported such issues. Participant responses on the speech, cello, and trombone tasks were also examined for sufficient variation across the 1-100 scale. Upon initial examination of the data, it became apparent that some participants were responding with either 1 or 100, but not making use of the scale in between. In some cases, it was believed they didn’t fully understand the instructions, but in others, participants likely just wanted to finish the study sooner. For the sound stimuli falling in the middle of the spectrum, such responses could certainly act as outliers, making data interpretation more complicated. Additionally, for the "matching" responses on the audiovisual tasks, response data were examined for similar anomalies like participants always pressing the same key. However, no exclusions were made based on these responses. Finally, data from a few participants were not used for analyses because collection was incomplete due to technical difficulties.

Predictor variables. One goal of this study was to investigate which types of music experience might influence a cognitive task such as audiovisual integration for speech or music perception (hypotheses 2 and 3). As such, a variety of predictors were selected that fit under a few categories. Some musical experiences were categorized as relating more to music "production" – that is, they were associated more with the act of making music. These variables included length of musical study (in years), type of instrument played, frequency of practice (hours per week), and number of performances given per month. Other variables were selected as being representative of more "receptive" musical activities – behaviors tied to perceiving music, but not necessarily creating it. These variables included time spent listening to music (hours per week), time spent watching digital musical performances like recorded concerts or music videos (hours per week), and number of live concerts attended per year. Additionally, PROMs melody and tuning task scores were used as easily quantifiable measures of certain music perception skills that could serve as predictors for everyone. However, these scores were expected to be correlated with the "production" music experience measures. "Type of instrument played" was coded for analysis given the categorical nature. Based on theoretical hypotheses regarding similarities and differences in instrument mechanics, participants were grouped into cellists (1), other string players (2), other instruments (3), and non-musicians (4) for the cello task, and trombonists (1), other brass players (2), other instruments (3), and non-musicians (4).

Dependent variables. For the main analyses of this study, dependent variables (DVs) primarily came in two forms. First, to assess visual influence across the whole sample (hypothesis 1), we compared mean sound stimulus ratings (for 5 sounds) on a scale from 1-100 based on their pairings with videos representing one extreme or the other of the scale (so two possible videos). For the speech task, these DVs were then ratings on five sounds ranging from "Va" to "Ba." For the cello and trombone vibrato tasks, the DVs were five sounds ranging from no vibrato to strong vibrato. For the cello pluck/bow task, the five sounds ranged from a fully plucked sound to a fully bowed sound. Five T tests were carried out for each task comparing specific sound stimulus responses (DVs) based on video pairing (Group).

Second, the primary dependent variable of interest for the multiple linear regressions (3 per task) and MANOVA involving speech music experience is based on evaluating the extent to which visual stimuli influence judgments of auditory stimuli in individuals (so that effects of personal experience, hypotheses 2 and 3, can be evaluated). This was established by calculating the difference between participants' audio ratings of each type of audio stimulus when paired with the two different videos, resulting in a "difference score." For example, take the cello vibrato task, with a response scale ranging from no vibrato (1) to very strong vibrato (100). In general, it is expected that any sounds paired with a strong vibrato video would be rated more highly on this scale, so a mean of responses from a sound (say, Vib 2 which is a medium strength vibrato sound) paired with the no vibrato video would be subtracted from the mean of responses for the same sound paired with the strong vibrato video. Determining this DV helps to assess the relationship between the musical experience predictor variables and the extent to which visual stimuli influenced auditory judgments.

Of note is that the “matching” responses for all audiovisual tasks (in which participants were asked to rate, on a scale from 1-5, the extent to which the audio and the video seemed to belong together or not) were not used as dependent variables for our analyses. This measure was primarily in place to motivate participants to watch the various videos while rating the accompanying auditory stimuli in the audiovisual tasks. Without this, participants may have been more likely to close their eyes or look away from the screen to avoid distracting perceptual information during these tasks.

Data organization and variable entry. Participant questionnaire responses and PROMS scores were organized on excel for later entry into SPSS. Experimental results were collected and consolidated through E-Prime 2.0 and OpenSesame, and eye-tracking data were gathered through GazePoint analysis software. All statistical analyses were carried out using SPSS software.

There is some theoretical motivation for predictor entry method for the multiple linear regression analyses, but relevant past findings utilizing the breadth of our predictors are scarce. As such, duration of musical experience and PROMS musical skills scores were entered first into a model hierarchically/blockwise (as more established indicators of formal musical experience). Principal Component Analyses were also conducted to determine if any underlying broader components were driving relationships across these predictor variables (types of experience and PROMS scores) and dependent variables. In addition to these methods, a backward selection multiple linear regression was utilized as an exploratory analysis given the lack of data regarding some types of musical experience.

Eye-tracking. Due to the large amount of data added in an eye-tracking paradigm, eye-tracking data were only collected for the cello vibrato task (beyond the questionnaire and PROMS tasks used as predictors). 15 participants completed the eye-tracking portion of this study; three were part of both the main experiment data collection and analyses and the eye-tracking portion, but for logistical reasons, the other twelve only completed the questionnaire, PROMS tasks, and cello vibrato task (for eye-tracking) and thus could not be included for the main analyses.

For each video, "areas of interest" (AOI) were created around different parts of the video so time spent looking at relevant vs. irrelevant (to vibrato production) movements could be assessed. If you recall from the earlier discussion on eye movements and expertise in chess, experts may look "between" areas of relevant information to assess the relationship between items (Reingold et al., 2001). Based on this information and initial observations from early eye-tracking data collection, two AOIs were established – these can be seen in Figure 4. One AOI is directly over the left hand performing the vibrato (and thus is the most relevant visual information for that sound), and the other AOI covers the upper body of the instrument and fingerboard, since this lies between the two main points of action in the video (the left hand and the bow). Due to the way in which eye-tracking recordings had to be captured, points of gaze were manually (visually) compared to AOIs using the data collected to determine if they fell within the designated areas. See Figure 5 for a still frame of gaze location and time displayed over the task.

Figure 4. Still frame from cello vibrato stimuli with Areas of Interest superimposed.

Figure 5. Screen capture from eye-tracking analysis with gaze path and fixations.

Using the established AOIs, several DVs were determined: left hand gaze time (AOI 1 above) in total seconds across trials, total "relevant" gaze time (time spent fixating in AOI 1+2 above) in total seconds across trials, and fixations per second (within AOIs 1 and 2). Fixations/second was added as a DV based on the findings from Reingold et al. (2001) suggesting that experts may display fewer fixations (regardless of duration) than novices for visual tasks.

CHAPTER 4

RESULTS

Multiple analyses were conducted to assess a) the presence of visual influence on auditory judgments across all participants (hypothesis 1), b) how musical experience might have impacted the extent of such influence (hypotheses 2 and 3), and c) how eye-gaze patterns might relate to either music experience or auditory judgments (hypotheses 4 and 5). Regarding hypothesis 1, t tests used to compare ratings of sounds based on their video accompaniment will be discussed. For hypotheses 2 and 3, data are presented from a MANOVA, Principal Components Analysis (PCA), and multiple linear regressions to assess the role of music experience in speech and music perception. Regarding hypotheses 4 and 5, a small sample was used, but results from correlation analyses, a PCA, and a simple linear regression based on the PCA are presented.

For the main analyses (hypotheses 1-3), Data were collected for 121 participants, but after exclusions, a sample of 108 participants was used for statistical analyses. Data from nine participants were excluded from these analyses because they consistently only rated sounds as either 1 or 100, not making use of the entire scale. Data from four additional participants were not used because tasks crashed, rendering data collection incomplete. The mean age of participants in the main analysis sample (N = 108) was 22.29 years old (SD = 4.99) and ranged from 18 to 35 years old. Of the 108 participants, 51 were male, and 57 were female. 78 participants reported having some kind of musical training (whether formal lessons or self-taught), and 30 reported having no musical training (though most likely received some form of musical guidance in school as children). Descriptive statistics for the additional music experience predictor variables is provided in Table 2.

Table 2

Main sample descriptive statistics for music experience predictors

Predictor Variable	Mean	SD	Range
Length of music study	9.63 years	7.01	1-32
Practice frequency	5.83 hours/week	8.46	0-39
Music listening	19.24 hours/week	17.36	1-96
Digital music viewing	3.38 hours/week	5.81	0-40
Live concert attendance	5.63 per year	10.03	0-60
PROMS melody score	10.38	3.91	1-20 (out of 20)
PROMS tuning score	8.75	2.82	3-16 (out of 18)

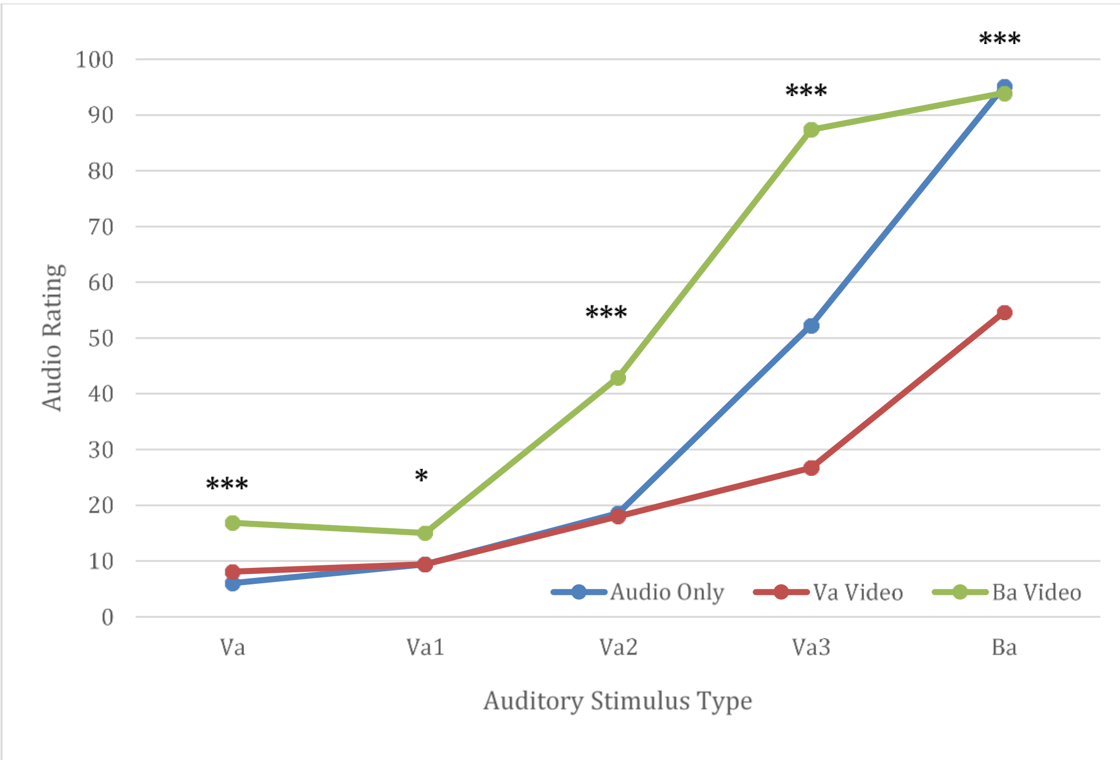
Note. Practice frequency and length of study only reflect means for participants with some music training.

General Visual Influence

To assess the effectiveness of the experimental manipulations and presence of general speech and music McGurk-like effects across all participants (hypothesis 1), we first compared audio ratings (dependent variable) for each sound based on its visual accompaniment (independent variable).

Speech task. For the speech task, this meant using t tests to compare sounds paired with the “Va” video to sounds paired with the “Ba” video. Auditory stimulus ratings based on video pairing are visualized across the intended sound spectrum in figure 6. Across all speech sounds, t tests suggested the differences in audio ratings (with 1 being fully /va/ and 100 being fully /ba/) based on video pairings were significant, suggesting a strong visual influence on auditory judgments. A table summarizing t test results comparing audio ratings based on video pairing can be seen in Table 3. These results remained significant when controlling for multiple comparisons. It should be noted that analyses for all speech sounds except one (Va 1) found

Levene's test of homogeneity of variances to be significant, suggesting unequal variances across groups based on video pairing. However, Brown-Forsythe analyses of these comparisons also found significant differences.



Note. * $p < 0.05$, *** $p < .001$.

Figure 6. Speech task audio ratings based on video accompaniment.

Table 3

T test results for speech sound ratings based on video accompaniment

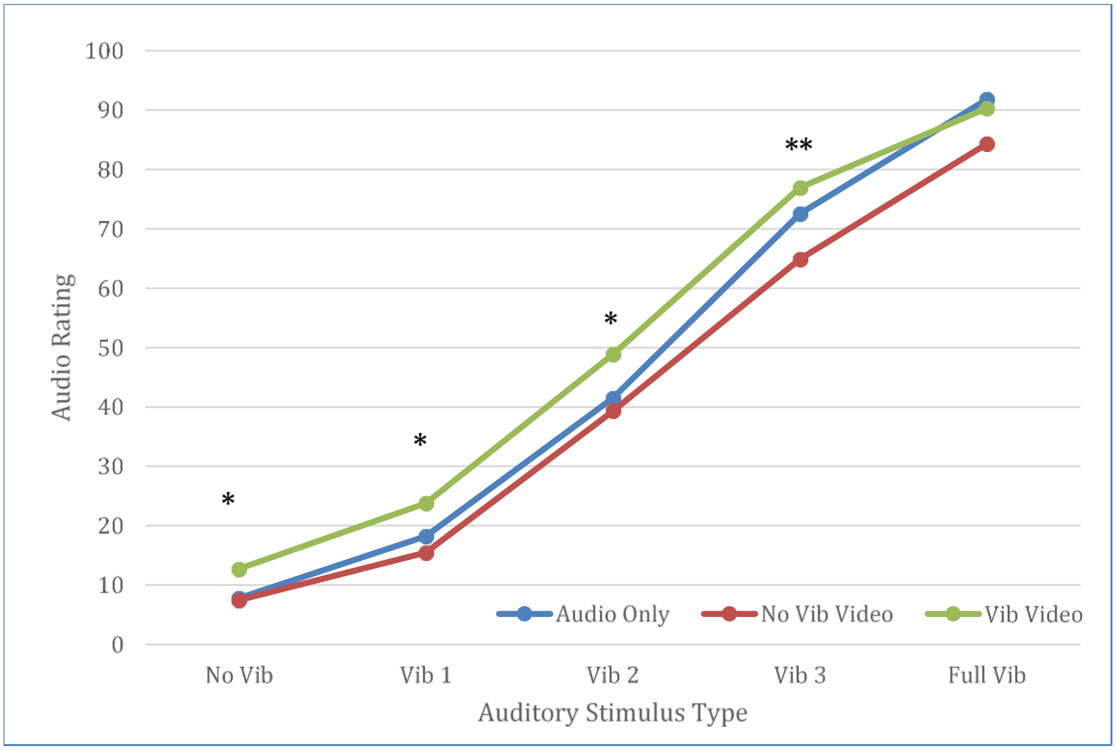
Sound	Group	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Va	Va Video	8.13	12.35	-4.03	214	.000**	.57
	Ba Video	17.18	18.78				
Va 1	Va Video	9.3	16.02	-2.24	214	.026*	.32
	Ba Video	14.31	16.82				
Va 2	Va Video	18.09	19.22	-7.45	214	.000**	1.01
	Ba Video	42.67	28.38				
Va 3	Va Video	27.73	24.48	-21.48	214	.000**	2.92
	Ba Video	88.43	16.23				
Ba	Va Video	56.91	34.1	-10.61	214	.000**	1.44
	Ba Video	93.74	11.78				

Note. *$p < .05$, **$p < .01$. Sound = auditory stimulus type, Group = video pairing type.

Cello vibrato task. For the cello vibrato task, t tests indicated several differences in audio ratings (where 1 = no vibrato and 100 = very strong vibrato) based on video pairing, though not to the same degree as the speech task. A line graph depicting the mean audio ratings based on video pairing across the auditory stimulus spectrum can be seen in Figure 7.

Audio ratings of the degree of vibrato heard differed significantly based on video pairing (videos of a note being played with and without vibrato) for the No Vib stimulus and the three more-ambiguous vibrato sounds in the middle of the spectrum, labeled Vib 1, Vib 2, and Vib 3. These differences remained significant when accounting for multiple comparisons. The No Vib analysis found a significant Levene's Test, suggesting a violation of the equal variances assumption, however a Brown-Forsythe analysis of this comparison also found a significant difference.

Table 4 summarizes results of a t test comparing cello vibrato sound ratings based on video pairing for these five sounds.



Note. * $p < 0.05$, ** $p < .01$

Figure 7. Cello vibrato ratings based on video accompaniment.

Table 4

T test results for cello vibrato sound ratings based on video accompaniment

Sound	Group	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
No Vib	No Vib Vid	6.46	13.46	-2.21	214	.028*	.3
	Vib Vid	11.02	16.92				
Vib 1	No Vib Vid	13.52	17.35	-3.51	214	.001**	.47
	Vib Vid	23.24	23.34				
Vib 2	No Vib Vid	38.66	27.08	-2.59	214	.01*	.35
	Vib Vid	47.97	26.34				
Vib 3	No Vib Vid	65.04	25.4	-3.64	214	.000**	.49
	Vib Vid	76.68	21.95				
Full Vib	No Vib Vid	85.47	18.65	-1.88	214	.062	.25
	Vib Vid	89.92	16.54				

Note. *$p < .05$. **$p < .0$. Sound = auditory stimulus type, Group = video pairing type.

Cello pluck/bow task. T tests were carried out for the cello pluck/bow task to compare responses to each sound (where 1 = a fully plucked sound and 100 = a fully bowed sound) when paired with either a pluck video or a bow video. Cello pluck/bow mean audio ratings based on video pairing across the auditory stimulus spectrum are presented in Figure 8. The cello pluck/bow revealed one significant difference in auditory judgments for the Full Pluck sound based on video pairing. A summary of the t test results can be seen in Table 5. The reduced number of significant differences based on video pairing across trials for this task deviates from the Saldaña and Rosenblum (1993) findings discussed earlier. The possible reasons for this are addressed in the discussion section.

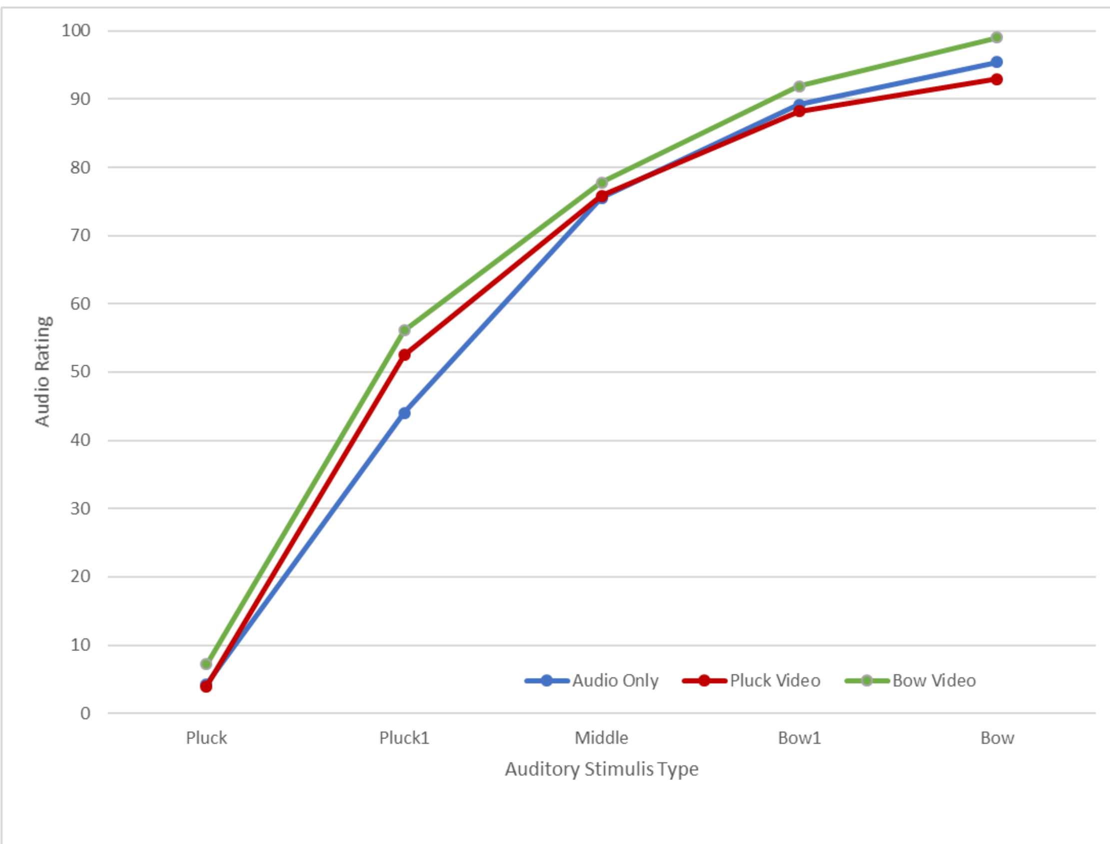


Figure 8. Cello pluck and bow audio ratings based on video accompaniment.

Table 5

T test results for cello pluck/bow sound ratings based on video accompaniment

Sound	Group	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Full Pluck	Pluck Vid	3.91	9.65	-2.065	214	.04*	.28
	Bow Vid	7.23	13.84				
Pluck 1	Pluck Vid	52.81	27.25	-.837	214	.404	.11
	Bow Vid	55.75	24.69				
Middle	Pluck Vid	76.79	18.55	.044	214	.965	.01
	Bow Vid	76.66	24.42				
Bow 1	Pluck Vid	87.93	15.61	-1.872	214	.062	.25
	Bow Vid	91.44	11.9				
Full Bow	Pluck Vid	94.06	12.6	-1.797	214	.074	.24
	Bow Vid	99.18	27.05				

Note. *$p < .05$, **$p < .01$. Sound = auditory stimulus type, Group = video pairing type.

Trombone vibrato task. Finally, regarding overall visual influence on the trombone task, t tests suggest that ratings for two sounds along the trombone vibrato spectrum (where 1 = no vibrato and 100 = very strong vibrato) may have been influenced significantly based on video pairing. A line graph depicting audio rating means based on video pairing across the stimulus spectrum can be seen in Figure 9. Specifically, the Vib 2 rating was significantly higher when paired with the vibrato video than when paired with the no vibrato video, and the Full Vib rating was significantly higher when paired with the vibrato video than when paired with the no vibrato video. However, when a False Discovery Rate (Banjamini & Hochberg, 1995) correction for multiple comparisons is considered, the difference for the Full Vib sound is no longer significant, so the strength of this finding is less clear. See Table 6 for a summary of these findings.

Note. * $p < 0.05$.

Figure 9. Trombone vibrato audio ratings based on video accompaniment.

Table 6

T test results for trombone vibrato sound ratings based on video accompaniment

Sound	Group	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
No Vib	No Vib Vid	5.55	14.41	-1.583	214	.115	.21
	Vib Vid	8.65	14.62				
Vib 1	No Vib Vid	18.38	19.84	-.246	214	.806	.03
	Vib Vid	19.05	20.7				
Vib 2	No Vib Vid	58.3	26.11	-2.233	214	.027*	.3
	Vib Vid	65.86	24.08				
Vib 3	No Vib Vid	70.9	23.65	-1.775	214	.077	.24
	Vib Vid	76.47	22.91				
Full Vib	No Vib Vid	80.48	21.14	-1.985	214	.048*	.27
	Vib Vid	85.71	17.81				

Note. *$p < .05$, **$p < .01$. Sound = auditory stimulus type, Group = video pairing type. Full Vib analysis no longer significant after controlling for multiple comparisons.

Taken together, the general visual influence findings suggest a strong speech McGurk effect, as well as a fairly strong musical McGurk effect in the case of vibrato. Presently there is weaker support for a musical McGurk effect for cello pluck/bow sounds and trombone vibrato sounds across all participants. These results partially indicate that the rejection of the null hypothesis for hypothesis 1 is appropriate. The weakness of the cello pluck/bow findings may be artifacts of the stimulus construction and will be examined in the discussion section.

Musical Experience

Predictor correlations. A bivariate correlation was conducted to assess relationships among music experience predictor variables. No evidence of multicollinearity was found, based on Pearson *r* values less than .8 (Field, 2009). See Table 7 for a summary of these results. Music “production” experiences like length of study and time spent practicing were found significantly positively related with one another, and with concert attendance and both PROMS tasks. The PROMS tasks were also significantly positively correlated with each other. Music listening was significantly positively correlated with both practice time and concert attendance.

Table 7

Music experience predictor variable Pearson correlations

	Prac	List	Cncrt	Dig	Mel	Tun
Lngth	.372**	-.077	.299**	-.095	.508**	.412**
Prac		.213*	.395**	.118	.235*	.378**
List			.211*	.403	-.063	.103
Cncrt				.118	.133	.153
Dig					-.203	-.010
Mel						.508**

Note. * $p < .05$. ** $p < .01$. Lngth = length of musical study in years. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week). Mel = PROMS melody score. Tun = PROMS tuning score.

Principal component analysis. A Principal Component Analysis (PCA) with oblique rotation was conducted to determine if various “musical experience” predictor variables could be organized and grouped on a theoretical basis. The Kaiser-Meyer-Olkin measure of sampling adequacy was sufficient ($KMO = .674$) based on criteria suggested by Field (2009). Bartlett’s test of sphericity, $\chi^2(21) = 142.624$, $p < .001$, indicated sufficient correlations between variables to conduct the PCA. “Production” music experience (duration of study and frequency of practice) loaded onto one component with the two PROMS scores, while the more “receptive” music activities like listening and digital viewing loaded onto a second component. Live concert attendance appeared to load onto both components, though somewhat weakly. Specific loadings for each predictor variable can be seen in Table 8. Largely these components make sense as distinct driving factors, especially considering that the PROMS scores were expected to (and did) correlate strongly with more production music experience (see correlations discussed above). However, subsequent analyses using these PCA components as predictors were largely not significant, and did not provide any stronger models than what individual variables could provide.

Table 8

Main study PCA pattern matrix loadings

Predictor	Component 1	Component 2
Length of Music Study (years)	.812	
Music Practice (hrs/week)	.599	
PROMS Melody Score	.766	
PROMS Tuning Score	.736	
Music Concert Attendance (per year)	.412	.446
Music Listening (hrs/week)		.814
Digital Music Viewing (hrs/week)		.743

Note. Loadings below .4 were suppressed (Field, 2009).

Instrument-specific experience.[1] A MANOVA was conducted to investigate differences in visual influence on auditory judgments based on what type of instrument participants played for each task (assessing hypotheses 2 and 3). Participants were grouped based on instrument type; there was a low but acceptable number of cellists (*N* = 10), but we felt there were not enough trombonists (*N* = 5) after participant exclusion to group participants based on trombone experience. These groups were compared on their "difference values;" the difference between participants' mean responses for each sound stimulus based on its pairing with one video or the other.

Results from these analyses were largely non-significant. There were no significant differences found for any speech difference scores based on type of instrument played – see Table B1 (in Appendix B) for a summary of this data. Similarly, no significant differences were found for any vibrato difference scores in the cello vibrato task – see Table B2 (in Appendix B)

[1] To aid in continuity, tables with data from entirely non-significant analyses can be found in Appendix B.

for a summary of this data. For one of the more ambiguous cello pluck/bow sounds (Bow 1), cellists (*M* = -8.32, *SD* = 9.12) were found to have a significantly smaller difference in audio ratings based on video pairing than other string players (*M* = 14.85, *SD* = 14.82) and non-musicians (*M* = 7.85, *SD* = 15.81). The results from the cello pluck/bow MANOVA can be seen in Table 9.

Table 9

MANOVA for cello pluck/bow difference scores based on instrument category

Sound	Instrument Group	*M*	*SD*	*F*	*df*	*p*	h_p^2
Pluck 1	Cellist	-10.04	11.19	.97	3, 100	.41	.027
	String player	1.59	11.24				
	Other musician	5.2	21.73				
	Non-musician	5.07	28.94				
Middle	Cellist	2.57	13.22	.75	3, 100	.527	.021
	String player	1.3	4.27				
	Other musician	-3.41	20.08				
	Non-musician	1.88	17.89				
Bow 1	Cellist	-8.32	9.12	5.236	3, 100	.002**	.131
	String player	14.85	14.82				
	Other musician	1.51	12.79				
	Non-musician	7.85	15.82				

Note. ** $p < .01$

For the trombone Vib 1 sound, "other musicians" (*M* = -3.13, *SD* = 20.08) – who were largely wind and brass players – had significantly smaller differences in audio ratings based on video pairing than non-musicians (*M* = 7.9, *SD* = 19.72). See table 10 for a summary of all the trombone MANOVA results. However, this finding is no longer significant when considering multiple comparisons using a False Discovery Rate correction (Banjamini & Hochberg, 1995).

Table 10

MANOVA for trombone vibrato difference scores based on instrument category

Sound	Instrument Group	*M*	*SD*	*F*	*df*	*p*	h_p^2
Vib 1	Cellist	6.44	10.54	2.7	3, 100	.049*	.071
	String player	-3.37	13.1				
	Other musician	-3.13	20.08				
	Non-musician	7.9	19.72				
Vib 2	Cellist	8.86	17.01	.044	3, 100	.987	.001
	String player	7.41	9.67				
	Other musician	7.95	21.45				
	Non-musician	6.56	19.56				
Vib 3	Cellist	-3.07	13.27	.918	3, 100	.435	.025
	String player	.33	13.48				
	Other musician	7.02	17.52				
	Non-musician	6.13	18.16				

Note. *$p < .05$.

While these findings allow for the possibility of effect of instrument type on difference scores, significant results were very limited, and interpretation of results from one sound type in the spectrum is difficult without a bigger pattern present. Taken more qualitatively however, these results suggest that if there is a difference based on instrument type, it may trend in the opposite direction from what was hypothesized. That is, musicians viewing their own instrument (or something more similar) may be less influenced by visual stimuli when making auditory judgments.

Evaluation of continuous music experience variables.[2] Multiple linear regression analyses were conducted to assess how the more continuous music experience variables might fit into models that predict differences between participants' auditory ratings based on video pairing across sounds and tasks (again, meant to assess hypotheses 2 and 3). To reduce the number of

[2] Tables with data from entirely non-significant analyses and exploratory analyses are located in Appendix B.

analyses run, it was decided to focus on the more ambiguous stimuli that fell along each spectrum (the "inner" three stimuli) since these were theorized to allow for greatest susceptibility to the influence of visual stimuli. Preliminary evaluations of the data suggested that linear models would be most appropriate based on data distributions. Broadly, the assumptions required for linear regression (linearity, multicollinearity, homoscedasticity, etc.) were met across analyses, though some analyses did not have normally distributed errors (several residual DV distributions were highly leptokurtic). Incidents of violation are addressed where relevant.

***Hierarchical multiple regression: blockwise variable entry*.** Multiple linear regressions were first run with a blockwise entry method to assess models for the various audio difference scores. Entry order was broadly motivated by past research suggesting processing differences based on "music production" experience, while less is known about "music reception" experience. Length of musical study was entered in the first block based on its frequent use as a measure of musical experience in past audiovisual music research (e.g. Paraskevopoulos et al., 2015; Hasegawa et al., 2004). The PROMS Melody and Tuning scores were also entered in the first block based on the past use of musical "aptitude" tasks in such research (e.g. Tiippana et al., 2013), as well as correlations with other music production variables in the present study. Frequency of music practice time (hrs/week) was added in a second block as a less often used variable also associated with music production experience. Music listening (hrs/week), live music viewing (concerts/year), and digital music viewing (hrs/week) were all entered in a third block as more exploratory measures associated with music reception experience to see if they explained variance beyond music production experience.

Using this entry method yielded significant regression models for only one difference score DV. All non-significant models can be found in Appendix B (Tables B3-B11). All models (model 1 used block 1, model 2 used blocks 1+2, and model 3 used blocks 1+2+3) were significant in predicting the trombone Vib 3 difference score; see Table 11 for the predictor correlations with the trombone Vib 3 DV, and Table 12 for a summary of the regression results. Within Model 1, length of musical study and the PROMS melody score were entered as negative coefficients, while the PROMS tuning score was entered as a positive coefficient. Model 2 (which added practice time) did explain additional variance meaningfully beyond what was explained by Model 1. Within Model 3, the PROMS melody score and live music viewing contributed as negative coefficients, while length of musical study, PROMS tuning, practice time, listening time, and digital music viewing all contributed as positive coefficients. While Model 1 offered a strong predictive model with only 3 variables, Model 3 explained about 6% more variance than Model 1, and had the greatest Adjusted R^2 value, suggesting it may provide the best fitting model. Of note, however, is that across all models, the PROMS melody score was the only "significant" coefficient to be included in a predictive model for the trombone Vib 3 difference scores, indicating it as a driving factor of these models.

Table 11

Pearson correlations between predictors and trombone Vib 3 difference scores

Predictor	R	Significance
Length of study (years)	-.146	.066
Practice time (hrs/week)	-.306	.001**
PROMS melody score	-.055	.287
PROMS tuning score	-.025	.401
Listening time (hrs/week)	.161	.048
Live music viewing (per year)	-.15	.062
Digital music viewing (hrs/week)	.080	.207

Note. ** $p < .01$.

Table 12

Hierarchical multiple linear regression for trombone Vib 3 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 change	Adjusted R^2	*F*	*df*	*p*
1	Lngth	-.023	.93	.326	.106	.106	.08	4.088	3, 104	.009**
	Mel	-1.668	.002							
	Tun	.782	.235							
2	Lngth	-.035	.899	.327	.107	.000	.072	3.043	4, 103	.02*
	Mel	-1.663	.003							
	Tun	.752	.273							
	Prac	.049	.871							
3	Lngth	.131	.644	.386	.149	.042	.089	2.48	7, 100	.022*
	Mel	-1.628	.003							
	Tun	.6	.382							
	Prac	.104	.745							
	List	.152	.166							
	Cnrct	-.308	.08							
	Dig	.087	.767							

Note. **p* < .05. ***p* < .01. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Exploration: backward selection. As a more exploratory approach, further multiple linear regressions were conducted to reassess the importance of our predictor variables to linear regression models predicting difference scores. In this case, variables were entered using the backward selection method (which starts by entering all predictors, then removes one and reassesses the model at each step) to avoid possible suppressor effects (Field, 2009). In this way, a backward selection approach is more likely (compared to forward selection) to detect if a specific variable is only important for a model in the presence of another variable. Again, no significant models were found for any of the speech sound difference scores, but some additional significant models (beyond what was found with blockwise entry) were found for the cello and trombone tasks. Full results from these analyses can be viewed in Appendix B, Tables B12-B23.

With the backward selection method, music experience variables were entered in significant models to predict the cello Pluck 1 sound, and all of the more ambiguous trombone vibrato (Vib 1, 2, and 3) sounds. In general, length of musical training and the PROMS melody variables were negatively related to the trombone difference scores, suggesting that greater music production experience and perception skills predicted smaller difference scores (and thus less visual influence of auditory judgments). Time spent practicing was the only music "production" variable to sometimes be positively associated with difference scores, and live concert attendance also varied based on the model and variable. Listening and digital music viewing were more consistently positively associated with difference scores.

The results of the multiple linear regression analyses across both blockwise entry and backward selection approaches do not reveal a strong overarching pattern among predictors or DVs, but offer some interesting observations. First, receptive music behaviors like listening to music and watching digital music performances were not expected to play much of a part in predicting the audio difference scores, but appear in models for a couple of analyses as being positively related to difference scores. When the music production variables (and PROMS melody score) do appear in models, they are most often negatively related to the difference scores (suggesting increased music ability associated with smaller difference scores), except in the case of music practice. Finally, while it is interesting that significant models were found for several more trombone task sounds than for cello sounds, the lack of consistency in predictors makes it unclear why or how this task might differ from the two cello tasks in its relationship to music experience.

The results of the MANOVA (investigating instrument-specific music experience) and the multiple linear regressions (assessing the continuous music experience variables) offer no evidence that the null hypothesis should be rejected for hypothesis 2. That is, no musical experience variables seemed to significantly impact the extent to which participants were influenced by visual stimuli in their auditory judgments for the speech task. There is some support for rejecting the null hypothesis for hypothesis 3, regarding the impact of musical experience (instrument-specific, music production, and music reception experience) on extent of visual influence in auditory judgments for the cello and trombone tasks; some music production and PROMS variables were found to predict difference scores. However, the results here are inconsistent. More research should be done to more strongly reject the null hypothesis (and make a case for music experience influencing difference scores on the cello and trombone tasks).

Eye-Tracking

The eye-tracking sample differed in a few important ways from the main experiment sample. The sample size for the eye-tracking portion of the study was much smaller ($N = 15$) than for the main analyses, making statistical analyses difficult to conduct. Data from two participants were excluded from the eye-tracking analyses due to technical difficulties with the equipment. The mean age was a bit older ($M = 30.3$ years old, $SD = 4.47$) and ranged from 23 to 40 years old, and there were proportionally more males ($N = 9$, 60%) than females ($N = 6$, 40%). Additionally, while there was good variation in musical experience among individuals reporting some form of musical study, only two participants (13%) reported having no musical training of any kind (formal or self-taught), compared to about 27% in the main sample. Descriptive statistics for the music experience predictor variables in this sample can be seen in Table 13.

Table 13

Eye-tracking sample descriptive statistics for music experience predictors

Predictor Variable	Mean	SD	Range
Length of music study	15.64 years	11.46	1-32
Practice frequency	5.83 hours/week	4.08	0-15
Performance frequency	2.27 per month	5.49	0-20
Music listening	15.64 hours/week	10.51	1-35
Digital music viewing	1.93 hours/week	3	0-10
Live concert attendance	11.21	14.74	0-55
PROMS melody score	13.23	3.9	7-19 (out of 20)
PROMS tuning score	9.15	2.94	3-14 (out of 18)

Bivariate Correlations were conducted to assess the strength of relationships between our sets of variables; broadly this meant relationships between musical experience variables and eye gaze variables, as well as eye gaze variables and task responses. Between music experience and eye gaze, we found that instrument category (string and non-string instruments) was significantly correlated (direction of correlation is relative to instrument category coding) with fixations per second within the AOIs (see results of t test below for further details here), that length of musical study was significantly positively correlated with amount of time viewing the cellist's left hand (AOI 1), and that the PROMS melody score was significantly positively correlated to left-hand gaze (AOI 1) and total gaze (AOI 1 + 2), and significantly negatively correlated with fixations per second. Significant Pearson correlation values are noted in Table 14. Regarding relationships between eye gaze and participant responses, only fixations per second were found to be significantly positively correlated with a difference score on the cello vibrato task (Vib 2), r = .559, p = .047.

Table 14

Music experience and eye gaze Pearson correlations

	List	Cncrt	Dig	Prac	Perf	Lngth	Mel	Tun	LHG	TG	Fix/Sec
Inst	-.049	-.407	.017	-.147	-.392	-.579*	-.724**	-.339	-.251	-.166	.588*
List		.557*	.399	.493	.687**	-.087	-.38	-.025	-.289	-.504	.133
Cncrt			.52	.756**	.958**	.304	.223	.07	-.07	.015	-.232
Dig				.609*	.478	.033	.304	.178	.108	.018	.151
Prac					.783**	.382	.456	.127	.101	.117	-.217
Perf						.405	.085	-.068	-.19	-.075	-.251
Lngth							.489	.317	.581*	.514	-.617*
Mel								.644*	.617*	.631*	-.569*
Tun									.466	.461	-.525
LHG										.615*	-.299
TG											-.535*

Note. $N = 15$, * $p < .05$. ** $p < .01$ Inst = instrument category. List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week). Prac = music practice (hrs/week). Perf = performances given per month. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. LHG = left-hand (AOI 1) gaze duration. TG = total AOI 1 + 2 gaze. Fix/Sec = fixations per second.

T tests were conducted to determine if the eye gaze DVs varied based on type of instrument played. Equal variances were assumed based on non-significant Levene's test results. An independent samples t test comparing individuals with string instrument experience ($N = 5$) and those with other instrument experience ($N = 8$) suggests that string players ($M = 1.45$, $SD = 0.55$) demonstrated fewer fixations per second in the AOIs than other instrument players ($M = 2.16$, $SD = 0.47$), $t(11) = -2.41$, $p = .034$. Figure 10 displays a stacked histogram representing the group distributions on this measure. This supports the findings discussed earlier by Reingold et al. (2001). No significant differences based on type of instrument played were found for the other gaze variables, and it should be noted that adding statistical corrections for multiple comparisons (Benjamini-Hochberg, 1995) resulted in no significant difference regarding fixations per second as well.

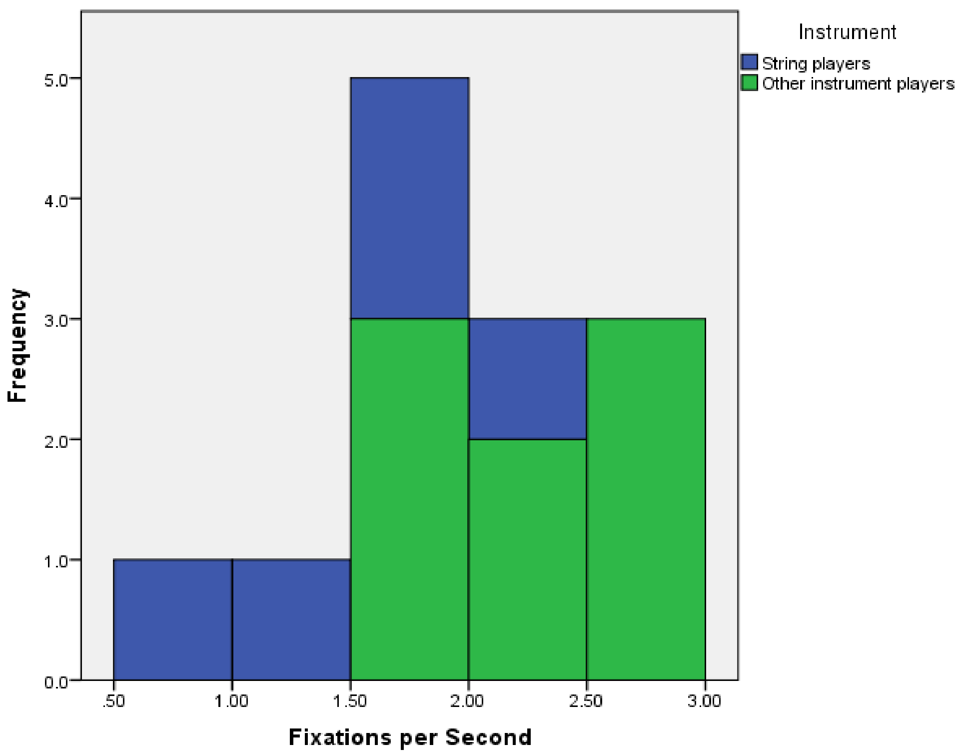


Figure 10. Histogram of fixation/second distribution by instrument type

A Principal Component Analysis with oblique rotation was conducted to determine if the musical experience predictor variables could be grouped on a theoretical basis for the eye-tracking portion of the study. The Kaiser-Meyer-Olkin measure of sampling adequacy was acceptable though low (*KMO* = .573) based on criteria suggested by Field (2009). Bartlett's test of sphericity, $\chi^2(28) = 68.52$, $p < .001$, indicated sufficient correlations between variables to conduct the PCA. Likely due to the small number of participants for the eye-tracking portion of the study, the components don't fully distinguish larger underlying driving factors, though

Component 1 contains the "receptive" musical activities (e.g. watching and listening) while component 2 includes scores on the PROMS tasks and years of musical study, possibly indicating stronger musical skills based on training. However, music practice and performances given (which are both more "production" experiences) both load more strongly onto component one, causing the ambiguity in interpretation. Table 15 indicates the relative loadings of each predictor in the PCA pattern matrix.

Table 15

Eye-tracking PCA pattern matrix loadings

Predictor	Component 1	Component 2
Music Listening (hrs/week)	.836	
Music Concert Attendance (per year)	.929	
Digital Music Viewing (hrs/week)	.907	
Music Practice (hrs/week)	.902	
Performances Given (per month)	.958	
Length of Musical Study (years)		.543
PROMS Melody Score		.919
PROMS Tuning Score		.803

Note. Loadings below .4 were suppressed (Field, 2009). The Structure Matrix was very similar to the Pattern Matrix and thus not reported here.

Two composite variables created from these components (by weighing individual predictors) were entered into a linear regression model to assess the extent to which they could be used to predict the amount of time spent viewing the cellist's left hand (AOI 1) during the cello vibrato task. Only component 2 was significantly correlated with left hand gaze time, $r(13)$

= .58, *p* = .023, and was the only one of the two variables entered into a significant regression model predicting left hand gaze, β = 32.146, *t*(11) = 2.66, *p* = .047. This component also explained a significant proportion of variance in left hand gaze times, R^2 = .34, *F*(1, 10) = 5.14, *p* =.047. Given the small *N* for the eye-tracking analyses, this analysis should be interpreted with caution, and no linear regressions are reported regarding the individual predictor variables.

Evaluation of hypotheses 4 and 5 (regarding the relationships between music experience and eye gaze, and eye gaze and extent of visual influence on auditory judgments) with these results is not recommended due to the small sample size. Taken as qualitative findings to guide future work, the results of the eye-tracking portion of the study overall more strongly suggest relationships between music experience (particularly instrument type and the PROMS scores) and gaze patterns than between gaze patterns and the extent to which participants' audio ratings were influenced by visual stimuli. The correlations, t tests, and PCA suggest that the PROMS melody score and type of instrument played may aid in predicting certain gaze patterns including left-hand gaze and fixations/second when watching a cellist play.

CHAPTER 5

DISCUSSION

This chapter first addresses the main study and exploratory eye-tracking study separately. For each, the research questions, hypotheses, and findings are reviewed, and implications and potential limitations are presented. The research (as a whole) is then discussed in the context of theories of speech and music perception, and future directions are suggested, followed by concluding remarks by the author.

Main Study

Questions, predictions, and findings. The research questions and aims for the main portion of the study broadly fall into two larger topics. First, is audiovisual integration for speech different in kind from audiovisual integration for other perceptual stimuli? This question was investigated by determining if visual articulatory cues influence the auditory perception (and categorization) of sounds similarly across speech and music presentations. Using a modified McGurk paradigm (see earlier discussions of methodology), ratings of various speech and musical instrument sounds were compared across video pairings with two different articulatory representations. It was predicted that on average across participants, the visual articulatory stimuli would significantly influence auditory judgments for both speech and music tasks.

Strong evidence of visual influence on auditory perception was found for speech sounds, but the evidence for such influence is less clear for the music tasks. The cello and trombone vibrato task results suggest that participants' audio ratings were influenced in the direction of the video (representing an end of the sound continuum) paired with the sound (e.g. Vib 2 paired with a no vib video), though these effects appear to be somewhat weaker and less consistent than for the speech task. Also, despite the results of the motivating study (Saldaña & Rosenblum, 1993), no significant differences in audio stimulus ratings were found based on video pairing in the cello pluck/bow task.

The second major question involves assessing the role of experience regarding visual sensory information influencing auditory perception. Is the extent to which we use visual information to inform auditory judgments dependent upon our experience with the stimulus? We used music as a context in which we could explore this question with a wide range of participant experiences. In this way, we hoped to assess what specific types of experience might matter in this context. We did this by observing how ratings for auditory stimuli varied within subjects depending on which visual stimulus accompanied the auditory stimulus, and determining if type of musical experience could help explain the degree of any differences found. It was predicted that musical experience would not help to predict these differences (in the form of "difference scores") for the speech tasks, but that music experience (particularly instrument-specific and production-based experiences) might predict greater susceptibility to visual influence for auditory judgments in a musical context.

The findings here offered nothing to suggest that musical experience could predict susceptibility to visual influence on a speech McGurk task; on the instrument-specific analysis (MANOVA), a theoretically-motivated multiple linear regression, and a more exploratory multiple linear regression, music experience did not seem to predict performance on the speech task in any significant way. However, the results for the music tasks are less conclusive. After accounting for multiple comparisons, only difference scores on one cello pluck/bow task varied depending on instrument-specific experience. Theoretically-motivated multiple linear regression analyses only found PROMS melody scores to predict smaller difference scores for one trombone stimulus, but results of an exploratory multiple linear regression, while still lacking some consistency, also found music production and PROMS melody variables to predict smaller music difference scores.

Implications and limitations. As with the Saldaña and Rosenblum (1993) study, results from the overall visual influence analyses (prior to accounting for music experience) partially support the existence of a significant "musical McGurk effect," however it generally appears weaker and less consistent than the effect for speech. Regarding the consistency, it is possible that due to the more continuous nature of the vibrato sound spectrum compared to the categorical nature of the pluck/bow sounds, the interpretation of ambiguous vibrato sounds is more susceptible to visual influence than the pluck/bow sounds. However, the speech sounds are more categorical in nature as well, and demonstrate a strong effect of visual influence on auditory ratings, so this seems a less likely explanation.

A couple of other factors may have impacted the consistency and strength of the music task results. First, creation of the pluck/bow stimuli was more difficult than creation of the vibrato stimuli, which were recorded naturally. In the Saldaña and Rosenblum (1993) study, the pluck/bow stimuli were created by taking a bowed sound, and chopping off more and more of the note onset sound until it sounded like a plucked note. However, for this study, it was felt that this method oversimplified the sound envelope differences from plucked and bowed sounds as they naturally occur. Thus, in addition to progressively shortening the note onset, sound features from both plucked and bowed sounds were layered to capture more nuances of note onset, sustain, and offset/decay. The result is that while participants overall rated these sounds in their intended order across the sound spectrum, the sound stimuli may have been more individually identifiable due to recognizable sound features from this editing process, and thus less susceptible to visual influence in the rating process. Additionally, a potential issue impacting consistency across all speech and music tasks is the possible lack of systematic variation across the sound spectrum for each sound type. While generally participants rated all sounds in their intended order from one end of the spectrum to the other, the steps between sounds were not necessarily equidistant, so that visual influence across an entire set of 5 sound stimuli for a task may not be evenly distributed or expected. Finally, the method of response to these stimuli may need modification as well. Our participants rated the sounds they heard on a scale from 1-100, but it may be that a smaller scale is more appropriate for this type of task, or that responding with a different method (e.g. sliding an indicator across the screen rather than entering numbers) may change the way participants categorize the stimuli and their subsequent responses. Finally, participants always responded regarding the extent of match/mismatch of the auditory and visual stimuli first, then rated the sounds they heard. This places slightly more demand on participant working memory when making their auditory rating judgment (possibly reducing the accuracy of ratings), and should be addressed in future studies.

A broader consideration that needs to be addressed when investigating possible differences between the speech and music tasks is the extent to which the speech and music sounds are comparable in terms of salient sound features and ambiguity. An important part of visual influence on auditory perception involves the extent to which information from each modality is thought to originate from the same source, allowing for multimodal binding. For making such judgments, some researchers have proposed a “unity assumption,” suggesting that below a certain threshold of asynchrony or mismatch in signals, individuals will still determine that two signals from different modalities came from the same event. However, other factors can influence where this threshold falls. A study by Chuen and Schutz (2016) first looked at how sensitive perceivers were to temporal offset of musical (cello and marimba) audiovisual stimuli. That is, how well did participants notice that the audio and visual stimuli were not in synch, and was their ability to notice mediated by the extent to which the perceived sound semantically "matched" the video or not? They found that participants were less likely to notice the audiovisual temporal offset when the sound and video "matched" (e.g. a cello video with a cello sound) than when they mismatched (e.g. a cello video with a marimba sound). By also manipulating the spectral information of each sound, they determined that both acoustic envelope (sound onset, sustain, offset) and spectral acoustic information (e.g. timbral sound features that allow you to tell the difference between a cello and a bassoon playing the same note) impacted the potential for audiovisual binding (which in turn influenced the sensitivity to temporal asynchrony). This is relevant to the present study in that some degree of audiovisual binding across our stimuli is likely necessary for visual influence of auditory judgments to take place. More simply put, the influence effect is probably going to be stronger if the pairing is somewhat believable. Subtleties of acoustic envelopes and spectral acoustic features that can influence this potential binding may differ across speech and music stimuli, so it may not just be that speech will always produce a stronger McGurk effect than music, but that appropriate stimuli are required to fully compare across domains.

Music experience. Regarding the analyses examining the role of music experience, conclusions are somewhat difficult to draw. One pattern to appear was that music experience never significantly predicted the extent to which participants' auditory ratings were influenced by visual stimuli on the speech task. This is at odds with the Proverbio et al. (2016) research that found musicians less susceptible to a speech McGurk illusion. However, the MANOVA and hierarchical multiple linear regression – combined with the exploratory regression analyses – hint at a potential pattern in which music production experience variables (like length of musical study and instrument played) and PROMS scores more consistently predicted smaller difference scores, offering the possibility that more "skilled" musicians were less swayed in their auditory judgments by visual stimuli. This is more in line with the assertion by Proverbio et al. that musicians may have more discriminant and finely tuned auditory perception systems, making them less susceptible (or reliant upon) visual cues. Our findings differ here however in that this ability seems more to depend on the context and domain of the stimuli.

A few factors may have impacted the strength and consistency of the findings regarding musical experience. First, much of the musical experience information (except for PROMS scores) for predictor variables was determined through self-report, leaving room for error in estimating things like frequency of practice, performance, etc. Second, and likely a bigger issue, is that the "musician" portion of our sample lacked the breadth of experience to really assess the role of musical experience in these tasks. A larger number of trombonists is necessary to properly assess the role of instrument-specific experience (on the trombone tasks) and identify a double dissociation (across musical tasks). Also, a wide range of musical experience was achieved between novices and graduate-level musicians, but ultimately lacked representation with true professional musicians due to the age range of the sample, and a lack of motivation to participate.

Eye-Tracking Study

Questions, predictions, and findings. If aspects of our audiovisual integration ability do seem to vary based on experience (as opposed to being universal and possibly innate across humans), could eye gaze patterns differ as a result of this experience, potentially facilitating any differences found in auditory judgments based on visual influence? We examined this question by determining if the music experience variables were systematically related to eye gaze location and fixations per second on our cello vibrato task, and if these eye-tracking variables were in turn related to task responses. It was predicted that participants with string instrument experience would look longer at the source of the vibrato (left hand) and display fewer fixations per second, and that longer gaze time at the left hand would predict greater differences in ratings for auditory stimuli based on their accompanying visual stimuli.

Despite a small sample size, music production experience (specifically instrument played and length of study) and the PROMS melody score tended to be significantly positively correlated with time spent in the relevant (to the task responses) areas of the screen, and negatively correlated with fixations per second in these areas (suggesting less jumping around). An independent samples t test also supported the possibility that string players had significantly fewer fixations per second in the relevant screen areas than non-string musicians. The only finding suggesting a relationship between eye gaze patterns and task responses was that fixations per second predicted (and were positively correlated with) difference scores on one vibrato stimulus.

Implications and limitations. The goal of the eye-tracking portion of the study was primarily to inform future research investigating eye-tracking and expertise, particularly in music. The sample size severely limits the extent to which we can meaningfully interpret the results or generalize them to the larger population. The accuracy and efficiency of the eye-tracking analyses also need to be improved from the current method in order to run a larger sample of participants with the equipment. The findings do, however, point in the direction of relationships between musical experience and some gaze patterns (identified above) during the cello vibrato task, which should be considered for future work. Based on its relationship with music production experience and PROMS predictor variables – and one of the difference scores for the cello vibrato task – fixations per second was identified as a potential important eye gaze variable that might serve to mediate or predict a relationship between musical skill and the extent of visual stimulus influence on auditory judgments in music tasks.

Implications for Theories of Speech and Auditory Perception

The design of this study was primarily established to assess certain claims of the Motor Theory of speech perception (Liberman & Mattingly, 1985), so what do the present findings offer toward the broader questions concerning the specialness of language sensory integration and the innateness of such processes? Our findings somewhat challenge the notion that speech is special for audiovisual integration by offering limited evidence of a musical McGurk effect, thus suggesting that the information from each modality is combined and processed in a similar manner across both speech and music. One could potentially make an argument that speech processing is special in the degree to which it utilizes the information from both modalities to inform the “intended articulatory gesture” (based on the strength of the overall effect in our speech task compared to the music tasks), but it seems more likely that the strength of the effect relies somewhat on the type of sound and visual articulation pairings presented (as it does for speech stimuli as well).

In assessing the potential innateness of speech processing abilities, our attempt to account for the influence of experience on such a process again resulted in a somewhat limited challenge of the Motor Theory due to inconsistent observed relationships between music experience and the degree to which participants demonstrated a sort of musical McGurk Effect. However, the relationships that were observed indicated that more extensive music production experience and perceptual "skill" may predict less reliance on visual cues when making auditory judgments, which is in line with the speech McGurk results utilizing native and foreign languages (Hayashi & Sekiyama, 1998) discussed earlier. This suggests then that certain strategies to processing and combining multimodal information are things that can be acquired through context-specific experience, rather than innate processes.

Though the present findings don't directly challenge the main perceptual claims made by the Motor Theory (e.g. that we perceive intended articulatory gestures made up of information from both modalities), it is the author's view that they can more easily fit other models of sensory perception that don't make such strong claims regarding specialization and innateness. The Direct-Realist theory of speech perception (Fowler, 1996), for example, would maintain that information about the physical gestures used to create a sound is perceived along with the sound itself (though not the "intended" gestures as Motor Theory suggests), but make no claim regarding specialized speech-specific or human-specific systems. Our findings don't clearly address this theoretical distinction between a necessity for motor-acoustic connections vs. acoustic signals alone being sufficient, though we did find some predictive differences between music production and music reception experience. Admittedly, there is likely a relationship between experience/familiarity with an instrument, and having strong visual-mechanical associations with the sound production of that instrument. Maes, Leman, Palmer, and Wanderley (2013) put forth a model of music perception that emphasizes a proposed overlap in brain representations for planning, execution, and perception of movement for music. Further, they argue that the integration of action and perception is based on associative learning (experience), indicating that a similar process for language (as the Motor Theory proposes) would neither be special for language, or specifically innate.

Interestingly, with the discovery of “mirror neurons” (as discussed in the literature review), some suggested a mechanism (or at the very least analogy) by which a Motor Theory of speech perception could be represented neurologically (Rizzolatti and Arbib, 1998). That is, when we perceive someone speaking, specific neurons in the brain associated with both producing and perceiving speech would activate. However, this doesn’t account for the way in which the auditory and motor articulations are proposed to combine into a coherent gesture. Further, mirror neurons contradict the other main arguments of the Motor Theory; mirror neurons were discovered in other primates (so not specific to humans), are used for other behaviors (so not specific to language), and proposed to be context- and experience-dependent to a certain degree, so while their existence may be innate, their functions are not (Lotto, Hickock, & Holt, 2009).

Other researchers (as summarized in Diehl, Lotto, and Holt, 2004) more broadly argue that auditory perception for speech is both non-specialized and also primarily concerned with processing of acoustic signals, rather than receiving information about a physical gesture along with the sound. For example, Kluender, Diehl, and Killeen (1987) found that Japanese quail could learn to categorize sounds based on syllable-initial sounds (regardless of vowels that followed), indicating that this ability is both a) not unique to humans and b) not reliant upon connections between our motor production abilities and the available acoustic signal (as the quail are physically and cognitively different in this respect from humans).

A more general auditory-only approach to sound perception could address our findings more simply here in terms of “a general ability of the perceiver to make use of multiple imperfect acoustic cues to categorize complex stimuli" (Diehl et al., 2004, p. 154). In this way, our participants are categorizing sounds with multiple (albeit sometimes ambiguous) acoustic cues based on cognitive representations of such sounds (whether from experience or training during the task, e.g. determining vibrato is "strong" when it is entirely relative), but information from the visual stimuli may be used to fill a void in missing acoustic information, or make a decision about incomplete acoustic information when categorizing. Thus, it may make more sense to address our findings in the context of the Fuzzy Logical Model of Perception (Oden & Massaro, 1978). This model utilizes the concept of "prototype" to suggest that for each category of stimulus (in this case speech and music sounds), we maintain a prototype against which we can compare new stimuli to decide how they fit. In this way, there is not a strict rule about necessary features, but more of a relative weighing of present and absent features relative to the prototype. While this model has been applied to speech perception, it acknowledges a broader use for auditory perception, and similar models have been postulated with categorization in other domains. It certainly seems reasonable then that in a situation with ambiguous auditory stimuli, we may look to other cues (even visual) to offer a valid comparison to our prototype. The main assertion here, however, is that our findings are more clearly and easily accounted for under alternative (and more domain-general) approaches to speech and sound perception compared to the Motor Theory.

Future Directions

Several directions to take this line of research are first recommended to address the previously-identified limitations of the study in its present form. Regarding the types of stimuli used, a few improvements can be made. Audiovisual speech and music stimuli should be evaluated more extensively to determine what might offer comparable and appropriate comparisons across domains for both modalities. This should be done both by examining the acoustic properties of sound stimuli used, and comparing features of the visual stimuli used such as range of motion associated with sound production. The extent to which stimuli for each modality map onto specific events in the accompanying modality should be assessed as well. For example, are the lip movements for a particular speech sound only associated with that sound, or would a speaker use the same (externally visible, at least) movements to create other sounds as well? Once selected, sound analysis software (e.g. a spectrogram) should be utilized to establish that speech and music stimuli are quantifiably varying in a systematic and equidistant manner.

To address issues of the representativeness of our sample, a concerted effort should be made to include professional musicians in the study. It may help to condense the study into its most useful parts (as it is currently long relative to the compensation available for non-student participants), and also construct an online version that participants can complete at home. With this approach, some experimental control is lost, but this may be worth it if a more appropriate and representative sample can be acquired. The eye-tracking component of the study (which needs many more participants) could not really be conducted under this methodology, but it could still be useful for the main analyses.

To approach some of our more general research questions, a few other directions are offered. First, it is impossible to definitively establish a causal relationship between music experience and the task responses under the current method, since music experience was not strictly “manipulated” as a variable for the study, but rather just observed. One alternative then would be to randomly assign (or match) participants to different types of musical (or other) training and see how such training might influence audiovisual integration in a musical context.

This would allow for both repeated-measures (pre- and post- training) analyses and group comparisons. While this approach would lose information regarding life-long musicians, it would gain experimental control.

Other extensions of our tasks could be implemented to address further influential factors. One possibility could involve presenting video stimuli that show more of the speaker's and musician's bodies to capture the relevant movements in a broader context. The present study intentionally focused on the smaller building blocks of music and speech production (both visually and auditorily), but the relevant motor movements often occur with other gestures and articulations (co-speech and co-music gestures) that, while not necessarily useful for sound production, may add to sound interpretation. In speech production and perception, hand gestures are used it disambiguate or emphasize auditory speech sounds, and research suggests that perceivers process "meaningful" hand gestures through brain connectivity between motor planning/production areas and semantic comprehension areas (Skipper, Goldin-Meadow, Nusbaum, & Small, 2009). This difference in processing compared to visual information like facial movements or non-meaningful hand gestures, suggests a unique function of these movements. In music, "we cannot see what a performer thinks but we can hear and see them move, and this may provide us with useful information about their conception of the music they are playing, or at least allow us to form an interpretation of what we think this conception might be" (Windsor, 2011, p. 48). If trying to assess the claim by the Motor Theory that we are perceiving "intended" gestures of speakers, then it seems important to account for other movements that may allow us to guess more strongly at what a sound producer's intention is.

An additional informative approach could involve utilizing scenarios (and perceptual conditions) that influence reliance on the information from both modalities. This could mean constructing the tasks to exist in a context of greater background noise (so that reliance on visual cues is increased), or it could mean running the same tasks with a sample of participants with partial hearing loss. Past research has indicated possible advantages for musicians for auditory perception in noisy conditions (Strait, Parbery-Clark, O'Connell, and Kraus, 2013), so running our speech task under such a condition might then yield more significant differences in reliance on visual stimuli based on music experience. An even more interesting manipulation could involve using a highly degraded or simplified visual signal paired with our auditory stimuli. One thought is to use motion-capture recording to simplify a musician's movements to moving dots on the screen. If the movements required to pluck or bow a cello string (or use vibrato) were simplified to such dots, and these dot movements were paired with our range of auditory stimuli, would the movements of the dots be salient enough to influence the auditory judgments of musicians (with intimate knowledge of the mechanics of that instrument), but not others?

Finally, to more strongly establish that audiovisual integration cognitive processes are more domain-general (as opposed to humans having a special speech sensory integration process), McGurk-like effects can be sought in other domains beyond speech and music. The researchers initially considered investigating such an effect in other contexts such as animal noises (e.g. a spectrum of dog and cat sounds paired with dog or cat images), but for simplicity and time constraints, this approach was not utilized for the present study.

Concluding Remarks

The primary aim in conducting this research was to investigate research questions regarding the generalizability and plasticity of human multisensory integration processes. However, the construction of the experiments – together with the results obtained – shed light on the complexity of the topics at hand. While we can account for varying types of musical behaviors, it remains difficult to define what it means to be a "musician." Further, the ways in which humans prioritize and attend to perceptual information in auditory and visual domains necessitates very particular methods of studying such phenomena.

The hope is that future iterations of such work can more definitively evaluate and inform theories of sensory integration and speech perception while accounting for the inherent complexity of the phenomena, but there is still much to understand about the nature of these processes. However, increased interest in these topics across disciplines like neuroscience, musicology, linguistics, and psychology will motivate collaborative and interdisciplinary approaches to tackle these issues. It is our wish that the present study exists as a bridge spanning these disciplines, and as a stepping stone for the continued exploration of human perception and cognition.

REFERENCES

Allman, B. L., & Meredith, M. A. (2007). Multisensory Processing in “Unimodal” Neurons: Cross-Modal Subthreshold Auditory Effects in Cat Extrastriate Visual Cortex. *Journal of Neurophysiology*, *98*(1), 545–549. http://doi.org/10.1152/jn.00173.2007

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual Integration of Speech Falters under High Attention Demands. *Current Biology, 15*(9), 839–843. http://doi.org/10.1016/j.cub.2005.03.046

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, *183*(3), 399–404. http://doi.org/10.1007/s00221-007-1110-1

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204–220. http://doi.org/10.1002/dev.20032

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of Audiovisual Integration Sites in Humans by Application of Electrophysiological Criteria to the BOLD Effect. *NeuroImage*, *14*(2), 427–438. http://doi.org/10.1006/nimg.2001.0812

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris, 98*(1-3), 191–205.

http://doi.org/10.1016/j.jphysparis.2004.03.018

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, *4*(1), 55-81.

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & cognition*, *29*(8), 1146-1152.

Chuen, L., & Schutz, M. (2016). The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude envelope cues. *Attention, Perception, & Psychophysics*, *78*(5), 1512-1528.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55*, 149-179.

Field, A. (2009). *Discovering statistics using SPSS*. Sage publications. Thousand Oaks, CA.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1730-1741.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523-552.

Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate behavioral research*, *26*(3), 499-510.

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception, & Psychophysics*, *77*(4), 1333-1341.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 172.

Hasegawa, T., Matsuki, K.-I., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., & Sadato, N. (2004). Learned audiovisual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cognitive Brain Research*, *20*(3), 510–518. http://doi.org/10.1016/j.cogbrainres.2004.04.005

Hayashi, Y., & Sekiyama, K. (1998). Native-foreign language effect in the McGurk effect: A test with Chinese and Japanese. In *AVSP'98 International Conference on Auditory-Visual Speech Processing*. Retrieved from http://www.isca-speech.org/archive_open/avsp98/av98_061.html

Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object Familiarity and Semantic Congruency Modulate Responses in Cortical Audiovisual Integration Areas. *Journal of Neuroscience*, *27*(30), 7881–7887. http://doi.org/10.1523/JNEUROSCI.1740-07.2007

Janthanasub, V., & Meesad, P. (2015). Evaluation of a low-cost eye tracking system for computer input. *King Mongkut's University of Technology North Bangkok International Journal of Applied Science and Technology*, *8*(3), 185-196.

Kaiser, J., Hertrich, I., Ackermann, H., Mathiak, K., & Lutzenberger, W. (2005). Hearing lips: gamma-band activity during audiovisual speech perception. *Cerebral Cortex*, *15*(5), 646-653.

Kayser, C., Lakatos, P., Meredith, A. (2012). Cellular physiology of cortical multisensory processing. In B. Stein (Ed.) *The New Handbook of Multisensory Processing* (115-134). Cambridge, MA: MIT Press.

Keil, J., Muller, N., Ihssen, N., & Weisz, N. (2011). On the variability of the McGurk effect: Audiovisual integration depends on prestimulus brain states. *Cerebral Cortex*, *22*(1), 221–231. http://doi.org/10.1093/cercor/bhr125

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, *237*(4819), 1195-1197.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*(3), 372–384. http://doi.org/10.1016/j.actpsy.2010.03.010

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences*, *105*(32), 11442–11445.

Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the Profile of Music Perception Skills. *PloS one*, *7*(12), e52508.

Lerdahl, F., & Jackendoff, R. (1983). A Generative Theory of Tonal Music. 1983. *to*, *352*, 346.

Lewkowicz, D. J. (2010). Infant perception of audiovisual speech synchrony. *Developmental Psychology*, *46*(1), 66–77. http://doi.org/10.1037/a0015579

Liberman, A., & Mattingly, I. (1985). The motor theory of speech perception revised. Retrieved from http://web.haskins.yale.edu/sr/SR082/SR082_06.pdf

Liu, B., Wang, Z., & Li, J. (2011). The influence of matching degrees of synchronous auditory and visual information in videos of real-world events on cognitive integration: an event-related potential study. *Neuroscience*, *194*, 19–26. http://doi.org/10.1016/j.neuroscience.2011.08.009

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in cognitive sciences*, *13*(3), 110-114.

Maes, P. J., Leman, M., Palmer, C., & Wanderley, M. (2014). Action-based effects on music perception. *Frontiers in psychology*, *4*, 1008.

Maravita, A., Bolognini, N., Bricolo, E., Marzi, C. A., & Savazzi, S. (2007). Is audiovisual integration subserved by the superior colliculus in humans? *Neuroreport*, *19*(3), 271–275.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. doi:10.3758/s13428-011-0168-7

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

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cognitive Brain Research*, *14*(1), 31–40.

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.

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: from motion to ’e’motion. *Social Cognitive and Affective Neuroscience*, *1*(3), 235–241. http://doi.org/10.1093/scan/nsl029

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*(40), 15894–15898.

Navarra, J., Vatakis, A., Zampini, M., Soto-Faraco, S., Humphreys, W., & Spence, C. (2005). Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cognitive Brain Research*, *25*(2), 499–507. http://doi.org/10.1016/j.cogbrainres.2005.07.009

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological review*, *85*(3), 172.

Paraskevopoulos, E., Kraneburg, A., Herholz, S. C., Bamidis, P. D., & Pantev, C. (2015). Musical expertise is related to altered functional connectivity during audiovisual integration. *Proceedings of the National Academy of Sciences*, *112*(40), 12522–12527. http://doi.org/10.1073/pnas.1510662112

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Musical Expertise Induces Audiovisual Integration of Abstract Congruency Rules. *Journal of Neuroscience*, *32*(50), 18196–18203. http://doi.org/10.1523/JNEUROSCI.1947-12.2012

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature neuroscience*, *6*(7), 674-681.

Petrini, K., McAleer, P., & Pollick, F. (2010). Audiovisual integration of emotional signals from music improvisation does not depend on temporal correspondence. *Brain Research*, *1323*, 139–148. http://doi.org/10.1016/j.brainres.2010.02.012

Proverbio, A. M., Calbi, M., Manfredi, M., & Zani, A. (2014). Audio-visuomotor processing in the Musician’s brain: an ERP study on professional violinists and clarinetists. *Scientific Reports*, *4*. http://doi.org/10.1038/srep05866

Proverbio, A. M., Massetti, G., Rizzi, E., & Zani, A. (2016). Skilled musicians are not subject to the McGurk effect. *Scientific reports*, *6*, 30423.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from https://www.pstnet.com

Recanzone, G. H. (2009). Interactions of auditory and visual stimuli in space and time. *Hearing Research*, *258*(1-2), 89–99. http://doi.org/10.1016/j.heares.2009.04.009

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in neurosciences*, *21*(5), 188-194.

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

Rucci, M., McGraw, P. V., & Krauzlis, R. J. (2016). Fixational eye movements and perception. *Vision research*, *118*, 1-4.

Ruth Clemo, H., Sharma, G. K., Allman, B. L., & Alex Meredith, M. (2008). Auditory projections to extrastriate visual cortex: connectional basis for multisensory processing in “unimodal” visual neurons. *Experimental Brain Research*, *191*(1), 37–47. http://doi.org/10.1007/s00221-008-1493-7

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, *54*(3), 406–416.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception, 36*(6), 888–897. http://doi.org/10.1068/p5635

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics*, *59*(1), 73–80.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends in Neurosciences*, *31*(8), 401–409. http://doi.org/10.1016/j.tins.2008.05.002

Senkowski, D., Molholm, S., Gomez-Ramirez, M., & Foxe, J. (2006). Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: a high-density electrical mapping study. *Cerebral Cortex, 16*(11), 1556-1565.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Current Biology*, *19*(8), 661-667.

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition*, *92*(3), B13–B23. http://doi.org/10.1016/j.cognition.2003.10.005

Stein, B. E. (Ed.) (2012). *The new handbook of multisensory processing. MIT Press.* Cambridge, MA.

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–1223. http://doi.org/10.1016/j.neuroimage.2008.09.034

Strait, D. L., Parbery-Clark, A., O'Connell, S., & Kraus, N. (2013). Biological impact of preschool music classes on processing speech in noise. *Developmental cognitive neuroscience*, *6*, 51-60.

Suied, C., Bonneel, N., & Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research*, *194*(1), 91–102. http://doi.org/10.1007/s00221-008-1672-6

Thompson, W., Graham, P., & Russo, F. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, *156*, 203-227.

Tiippana, K., Viitanen, K., & Kivimäki, R. (2013). The effect of musical aptitude on the integration of audiovisual speech and non-speech signals in children. In *AVSP* (pp. 153–156). Retrieved from http://avsp2013.loria.fr/proceedings/papers/paper_23.pdf

Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and Non-Speech Audiovisual Illusions: A Developmental Study. *PLoS ONE*, *2*(8), e742. http://doi.org/10.1371/journal.pone.0000742

von Stein, A., Rappelsberger, P., Sarnthein, J., & Petsche, H. (1999). Synchronization between temporal and parietal cortex during multimodal object processing in man. *Cerebral Cortex*, *9*(2), 137-150.

Windsor, W. L. (2011). Gestures in music-making: Action, information and perception. In A. Gritten (Ed.), New Perspectives on Music and Gesture, Chapter 2, pp. 45–66. Farnham, UK: Ashgate.

Yi, A., Wong, W., & Eizenman, M. (2013). Gaze patterns and audiovisual speech enhancement. *Journal of Speech, Language, and Hearing Research*, *56*(2), 471-480.

APPENDICES

APPENDIX A

Background Questionnaire Questions

General Information

Gender _______

What is your age in years? __________

Have you ever been diagnosed with a hearing impairment? Yes / No

Have you ever been diagnosed with an attention or learning disorder? Yes / No
If so, which one(s)? ________________________________

Do you consider English your native language? Yes / No

Do you currently, or have you ever, used more than one language in day-to-day life?
If so, what languages? _________________________

Musician Experience

How many hours in a week (on average) do you spend listening to music (not counting while also playing music)? _____________
What types/genres of music do you listen to? _______________

How many times a month do you *watch* musicians perform digitally (e.g. recorded concerts, music videos) (e.g. attend live concerts, watch music videos, etc.)? ________________

How many times a month do you *watch* musicians perform live (e.g. attend live concerts)? _______________

Do you currently, or have you ever, played any musical instruments? Yes / No
If so, what instrument(s)? ________________________

At what age did you start formal training (please list for all instruments played)? __________

How many years of formal training have you had (please list for all instruments played)?

How many hours do you currently practice or rehearse your instrument(s) per week?

How many times a month do you perform in a concert setting (solo or ensemble)?

APPENDIX B

Supplemental Statistical Analysis Tables

Table B1

MANOVA for speech difference scores based on instrument category

Sound	Instrument Group	*M*	*SD*	*F*	*df*	*p*	h_p^2
Va 1	Cellist	8.32	10.19	.479	3, 100	.698	.014
	String player	3.41	10.56				
	Other musician	3.19	25.18				
	Non-musician	8.33	15.27				
Va 2	Cellist	13.15	13.15	.706	3, 100	.551	.02
	String player	27.92	32.05				
	Other musician	27.62	33.67				
	Non-musician	20.19	31.39				
Va 3	Cellist	65.09	25.71	.185	3, 100	.907	.005
	String player	54.33	31.8				
	Other musician	61.39	29.8				
	Non-musician	60.18	34.45				

Table B2

MANOVA for cello vibrato difference scores based on instrument category

Sound	Instrument Group	*M*	*SD*	*F*	*df*	*p*	h_p^2
Vib 1	Cellist	15.64	5.31	1.645	3, 100	.184	.045
	String player	-.41	15.92				
	Other musician	6.97	17.21				
	Non-musician	11.61	19.83				
Vib 2	Cellist	6.11	11.42	.325	3, 100	.807	.009
	String player	8	24.45				
	Other musician	7.92	26.27				
	Non-musician	13.01	25.95				
Vib 3	Cellist	.89	21.19	.834	3, 100	.478	.023
	String player	5.22	13.67				
	Other musician	12.77	22.77				
	Non-musician	13.16	25.72				

Table B3

Hierarchical multiple linear regression for speech Va 1 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	Lngth	-.181	.592	.069	.005	.005	-.024	.165	3, 104	.920
	Mel	-.067	.922							
	Tun	.440	.6							
2	Lngth	-.071	.839	.136	.018	.014	-.02	.481	4, 103	.750
	Mel	-.115	.866							
	Tun	.719	.408							
	Prac	-.453	.235							
3	Lngth	-.034	.924	.231	.054	.035	-.013	.8	7, 100	.589
	Mel	-.221	.746							
	Tun	.869	.321							
	Prac	-.397	.331							
	List	-.179	.197							
	Cnrct	-.078	.727							
	Dig	.648	.087							

Note. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Table B4

Hierarchical multiple linear regression for speech Va 2 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	Lngth	-.018	.972	.158	.025	.025	-.003	.882	3, 104	.453
	Mel	-1.042	.311							
	Tun	-.559	.658							
2	Lngth	-.003	.996	.159	.025	.000	-.013	.659	4, 103	.622
	Mel	-1.048	.311							
	Tun	-.52	.693							
	Prac	-.064	.912							
3	Lngth	.155	.778	.238	.057	.032	-.01	.851	7, 100	.548
	Mel	-.896	.388							
	Tun	-.767	.563							
	Prac	-.306	.621							
	List	.352	.097							
	Cnrct	-.024	.943							
	Dig	.003	.996							

Note. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Table B5

Hierarchical multiple linear regression for speech Va 3 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	Lngth Mel Tun	-.130 --1.021 1.227	.79 .302 .313	.131	.017	.017	-.011	.6	3, 104	.617
2	Lngth Mel Tun Prac	-.148 -1.013 1.182 .074	.771 .309 351 .894	.132	.017	.000	-.021	.45	4, 103	.772
3	Lngth Mel Tun Prac List Cnrct Dig	-.061 --.986 .987 .371 -.114 .443 -.642	908 .324 .439 .533 .575. .176 .244	.223	.05	.033	-.017	.742	7, 100	.637

Note. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Table B6

Hierarchical multiple linear regression for cello Vib 1 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	Lngth Mel Tun	-.186 .762 -1.030	.508 .181 .142	.178	.032	.032	.003	1.104	3, 104	.351
2	Lngth Mel Tun Prac	-.299 .815 -1.330 .472	.303 .151 .068 .136	.231	.053	.021	.015	3.043	4, 103	.238
3	Lngth Mel Tun Prac List Cnrct Dig	-.283 .806 -1.371 .611 -.001 -.166 -.219	.352 .160 .064 .075 .991 .372 .487	.261	.068	.015	.001	2.48	7, 100	.427

Note. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Table B7

Hierarchical multiple linear regression for cello Vib 2 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 change	Adjusted R^2	*F*	*df*	*p*
1	Lngth	-.369	.360	.094	.009	.009	-.02	.303	3, 104	.823
	Mel	.421	.605							
	Tun	-.065	.948							
2	Lngth	-.430	.307	.108	.012	.003	-.027	.301	4, 103	.877
	Mel	.447	.585							
	Tun	-.219	.834							
	Prac	.249	.586							
3	Lngth	-.580	.186	.192	.037	.025	-.031	.542	7, 100	.801
	Mel	.362	.661							
	Tun	-.046	.965							
	Prac	.364	.460							
	List	-.230	.171							
	Cnrct	.125	.643							
	Dig	-.116	.798							

Note. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Table B8

Hierarchical multiple linear regression for cello Vib 3 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 change	Adjusted R^2	*F*	*df*	*p*
1	Lngth	-.484	.185	.14	.019	.019	-.009	.681	3, 104	.565
	Mel	.658	.372							
	Tun	-.191	.832							
2	Lngth	-.514	.177	.143	.02	.001	-.018	.529	4, 103	.715
	Mel	.671	.365							
	Tun	-.267	.777							
	Prac	.123	.766							
3	Lngth	-.462	.238	.259	.067	.047	.001	1.017	7, 100	.424
	Mel	.562	.446							
	Tun	-.091	.923							
	Prac	.103	.814							
	List	-.181	.227							
	Cnrct	-.022	.927							
	Dig	.883	.032							

Note. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Table B9

Hierarchical multiple linear regression for cello Pluck 1 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	Lngth Mel Tun	-.674 -.526 1.638	.078 .493 .085	.236	.056	.056	.028	2.021	3, 104	.116
2	Lngth Mel Tun Prac	-.585 -.565 1.863 -.366	.140 .463 .060 .395	.25	.062	.007	.026	1.694	4, 103	.157
3	Lngth Mel Tun Prac List Cnrct Dig	-.420 -.434 1.636 -.572 .319 -.065 .071	.301 .571 .097 .212 .042 .793 .867	.331	.109	.047	.046	1.738	7, 100	.109

Note. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Table B10

Hierarchical multiple linear regression for cello pb Middle difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	Lngth Mel Tun	-.113 -.356 1.214	.688 .535 .089	.169	.028	.028	000	.985	3, 104	.403
2	Lngth Mel Tun Prac	-.043 -.387 1.396 -.291	.884 .500 .060 .360	.191	.037	.008	-.002	.949	4, 103	.439
3	Lngth Mel Tun Prac List Cnrct Dig	-.099 -.398 1.419 -.239 -.051 .033 -.184	.746 .495 .061 .489 .666 .860 .564	.211	.045	.008	-.024	.647	7, 100	.716

Note. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Table B11

Hierarchical multiple linear regression for cello Bow 1 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	Lngth	-.220	.344	.132	.017	.017	-.011	.606	3, 104	.612
	Mel	.474	.312							
	Tun	-.408	.479							
2	Lngth	-.149	.534	.171	.029	.012	-.009	.765	4, 103	.551
	Mel	.443	.345							
	Tun	-.229	.700							
	Prac	-.289	.269							
3	Lngth	-.129	.603	.271	.074	.044	.008	1.122	7, 100	.355
	Mel	.496	.288							
	Tun	-.256	.668							
	Prac	-.499	.075							
	List	.099	.297							
	Cnrct	.171	.263							
	Dig	.201	.434							

Note. Lngth = length of musical study in years. Mel = PROMS melody score. Tun = PROMS tuning score. Prac = music practice (hrs/week). List = music listening time (hrs/week). Cncrt = music concert attendance per year. Dig = digital music watching (hrs/week).

Table B12

Backward multiple linear regression for speech Va 1 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	MusicDuration PROMSmelody PROMStuning MaxPractice listening live digital	-.034 -.221 .869 -.397 -.179 -.078 .648	.924 .746 .321 .331 .197 .727 .087	.231	.054	.054	-.013	.8	7, 100	.589
2	PROMSmelody PROMStuning MaxPractice listening live digital	-.246 .859 -.405 -.178 -.082 .652	.696 .320 .307 .196 .705 .083	.231	.053	.000	-.003	.941	6, 101	.489
3	PROMSmelody PROMStuning MaxPractice listening digital	-.262 .870 -.456 -.184 .648	.675 .312 .220 .175 .083	.228	.052	-.001	.005	1.11	5, 102	.36
4	PROMStuning MaxPractice listening digital	.704 -.469 -.176 .646	.354 .204 .188 .083	.225	.05	-.002	.013	1.344	4, 103	.255
5	MaxPractice listening digital	-.342 -.170 .619	.317 .201 .095	.206	.042	-.008	.014	1.519	3, 104	.214
6	listening digital	-.195 .606	.138 .102	.182	.033	-.009	.014	1.774	2, 105	.175
7	digital	.385	.257	.11	.012	-.021	.003	1.297	1, 106	.257

Note. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B13

Backward multiple linear regression for speech Va 2 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 change	Adjusted R^2	*F*	*df*	*p*
1	MusicDuration	.155	.778	.238	.057	.057	-.010	.851	7, 100	.548
	PROMSmelody	-.896	.388							
	PROMStuning	-.767	.563							
	MaxPractice	-.306	.621							
	listening	.352	.097							
	live	-.024	.943							
	digital	.003	.996							
2	MusicDuration	.155	.776	.238	.057	.000	.000	1.003	6, 101	.428
	PROMSmelody	-.896	.385							
	PROMStuning	-.767	.560							
	MaxPractice	-.305	.619							
	listening	.352	.073							
	live	-.024	.943							
3	MusicDuration	.147	.782	.238	.057	.000	.010	1.214	5, 102	.308
	PROMSmelody	-.895	.383							
	PROMStuning	-.761	.561							
	MaxPractice	-.318	.588							
	listening	.350	.069							
4	PROMSmelody	-.787	.405	.237	.056	-.001	.019	1.512	4, 103	.204
	PROMStuning	-.716	.579							
	MaxPractice	-.271	.628							
	listening	.342	.071							
5	PROMSmelody	-.824	.380	.232	.054	-.002	.026	1.952	3, 104	.126
	PROMStuning	-.896	.468							
	listening	.323	.080							
6	PROMSmelody	-1.169	.148	.221	.049	-.005	.031	2.674	2, 105	.074
	listening	.303	.096							
7	listening	. .323	.077	.172	.030	-.019	.020	3.194	1, 106	.077

Note. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B14

Backward multiple linear regression for speech Va 3 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 change	Adjusted R^2	*F*	*df*	*p*
1	MusicDuration PROMSmelody PROMStuning MaxPractice listening live digital	-.061 -.986 .987 .371 .114 -.443 -.642	.908 .324 .439 .533 .575 .176 .244	.223	.050	.050	-.017	.742	7, 100	.637
2	PROMSmelody PROMStuning MaxPractice listening live digital	-1.03 .968 .356 .117 -.451 -.636	.263 .442 .537 .559 .157 .244	.223	.050	.000	-.007	.872	6, 101	.518
3	PROMSmelody PROMStuning MaxPractice live digital	-1.11 1.049 .393 -.427 -.514	.220 .401 .492 .174 .306	.216	.046	-.003	-.001	.984	5, 102	.431
4	PROMSmelody PROMStuning live digital	-1.09 1.307 -.350 -.480	.228 .272 .231 .335	.205	.042	-.004	.004	1.117	4, 103	.353
5	PROMSmelody PROMStuning live	-1.05 1.318 -.384	.243 .267 .186	.182	.033	-.009	.005	1.178	3, 104	.322
6	PROMSmelody live	-.565 -.353	.473 .222	.147	.021	-.012	.003	1.142	2, 105	.323
7	live	-.379	.186	.129	.017	-.005	.007	1.773	1, 106	.186

Note. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B15

Backward multiple linear regression for cello Vib 1 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 change	Adjusted R^2	*F*	*df*	*p*
1	MusicDuration	-.283	.352	.261	.068	.068	.001	1.013	7, 100	.427
	PROMSmelody	.806	.160							
	PROMStuning	-1.371	.064							
	MaxPractice	.611	.075							
	listening	-.001	.991							
	live	-.166	.372							
	digital	-.219	.487							
2	MusicDuration	-.282	.345	.261	.068	.000	.011	1.194	6, 101	.316
	PROMSmelody	.807	.156							
	PROMStuning	-1.372	.060							
	MaxPractice	.611	.071							
	live	-.167	.363							
	digital	-.220	.450							
3	MusicDuration	-.246	.675	.250	.063	-.005	.015	1.324	5, 102	.26
	PROMSmelody	.798	.312							
	PROMStuning	-1.362	.220							
	MaxPractice	.579	.175							
	live	-.181	.083							
4	PROMSmelody	.610	.240	.237	.056	-.007	.018	1.483	4, 103	.213
	PROMStuning	-1.433	.048							
	MaxPractice	.528	.108							
	live	-.208	.244							
5	PROMSmelody	.582	.262	.207	.043	-.013	.015	1.513	3, 104	.216
	PROMStuning	-1.413	.051							
	MaxPractice	.388	.204							
6	PROMStuning	-1.041	.105	.176	.031	-.012	.012	1.629	2, 105	.201
	MaxPractice	.403	.187							
7	PROMStuning	-.718	.227	.119	.014	-.017	.005	1.479	1, 106	.227

Note. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B16

Backward multiple linear regression for cello Vib 2 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	MusicDuration PROMSmelody PROMStuning MaxPractice listening live digital	-.580 .362 -.046 .364 -.230 .125 -.116	.186 .661 .965 .460 .171 .643 .798	.192	.037	.037	-.031	.542	7, 100	.801
2	MusicDuration PROMSmelody MaxPractice listening live digital	-.583 .348 .358 -.231 .125 -.115	.178 .647 .450 .164 .639 .798	.192	.037	.000	-.021	.638	6, 101	.699
3	MusicDuration PROMSmelody MaxPractice listening live	-.571 .343 .352 -.246 .122	.182 .650 .455 .111 .645	.190	.036	-.001	-.011	.760	5, 102	.581
4	MusicDuration MaxPractice listening live	-.482 .366 -.249 .117	.202 .435 .105 .658	.185	.034	-.002	-.004	.905	4, 103	.464
5	MusicDuration MaxPractice listening	-.448 .423 -.237	.224 .347 .115	.180	.032	-.002	.004	1.15	3, 104	.333
6	MusicDuration listening	-.308 -.199	.360 .169	.155	.024	-.008	.005	1.279	2, 105	.283
7	listening	-.187	.193	.127	.016	-.008	.007	1.715	1, 106	.193

Note. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B17

Backward multiple linear regression for cello Vib 3 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 change	Adjusted R^2	*F*	*df*	*p*
1	MusicDuration PROMSmelody PROMStuning MaxPractice listening live digital	-.462 .562 -.091 .103 -.181 -.022 .883	.238 .446 .923 .814 .227 .927 .032	.259	.067	.067	.001	1.017	7, 100	.424
2	MusicDuration PROMSmelody PROMStuning MaxPractice listening digital	-.469 .564 -.086 .092 -.184 .882	.218 .443 .927 .826 .214 .031	.259	.067	.000	.011	1.197	6, 101	.314
3	MusicDuration PROMSmelody MaxPractice listening digital	-.473 .538 .083 -.185 .884	.209 .426 .838 .205 .029	.259	.067	.000	.021	1.449	5, 102	.213
4	MusicDuration PROMSmelody listening digital	-.448 .546 -.178 .890	.205 .417 .208 .027	.258	.067	.000	.030	1.818	4, 103	.131
5	MusicDuration listening digital	-.301 -.183 .900	.320 .194 .025	.246	.060	-.006	.033	2.21	3, 104	.091
6	listening digital	-.176 .928	.212 .021	.227	.051	-.009	.033	2.815	2, 105	.064
7	digital	.728	.047	.192	.037	-.014	.028	4.035	1, 106	.047*

Note. *Model 7 not significant after multiple comparisons correction. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B18

Backward multiple linear regression for cello Pluck 1 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	MusicDuration PROMSmelody PROMStuning MaxPractice listening live digital	-.420 -.434 1.636 -.572 .319 -.065 .071	.301 .571 .097 .212 .042 .793 .867	.331	.109	.109	.046	1.738	7, 100	.109
2	MusicDuration PROMSmelody PROMStuning MaxPractice Listening live	-.427 -.427 1.626 -.567 .329 -.064	.290 .575 .097 .213 .024 .798	.330	.109	.000	.056	2.043	6, 101	.067
3	MusicDuration PROMSmelody PROMStuning MaxPractice listening	-.448 -.424 1.641 -.600 .322	.254 .576 .092 .168 .024	.330	.109	-.001	.064	2.461	5, 102	.038*
4	MusicDuration PROMStuning MaxPractice listening	-.531 1.438 -.590 .329	.143 .109 .173 .021	.325	.106	-.003	.071	3.017	4, 103	.021*
5	MusicDuration PROMStuning listening	-.678 1.145 .283	.052 .190 .041	.299	.089	-.017	.063	3.366	3, 104	.022*
6	MusicDuration listening	-.490 .308	.123 .025	.272	.074	-.015	.056	4.149	2, 105	.018*
7	listening	.327	.018	.229	.052	-.022	.043	5.8	1, 106	.018*

Note. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B19

Backward multiple linear regression for cello pb Middle difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	MusicDuration	-.099	.746	.211	.045	.045	-.024	.647	7, 100	.716
	PROMSmelody	-.398	.495							
	PROMStuning	1.419	.061							
	MaxPractice	-.239	.489							
	listening	-.051	.666							
	live	.033	.860							
	digital	-.184	.564							
2	MusicDuration	-.088	.768	.210	.044	.000	-.014	.757	6, 101	.606
	PROMSmelody	-.400	.491							
	PROMStuning	1.410	.061							
	MaxPractice	-.222	.501							
	listening	-.048	.680							
	digital	-.182	.567							
3	PROMSmelody	-.466	.384	.208	.043	-.001	-.005	.899	5, 102	.485
	PROMStuning	1.385	.063							
	MaxPractice	-.251	.425							
	listening	-.044	.701							
	digital	-.174	.581							
4	PROMSmelody	-.436	.407	.205	.042	-.001	.004	1.096	4, 103	.363
	PROMStuning	1.358	.065							
	MaxPractice	-.271	.380							
	digital	-.220	.446							
5	PROMSmelody	-.421	.423	.191	.036	-.006	.008	1.271	3, 104	.289
	PROMStuning	1.384	.060							
	MaxPractice	-.303	.321							
6	PROMStuning	1.114	.086	.174	.030	-.006	.011	1.588	2, 105	.209
	MaxPractice	-.316	.300							
7	PROMStuning	.86	.151	.141	.020	-.010	.010	2.089	1, 106	.151

Note. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B20

Backward multiple linear regression for cello Bow 1 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	MusicDuration PROMSmelody PROMStuning MaxPractice listening live digital	-.129 .496 -.256 -.499 .099 .171 .201	.603 .288 .668 .075 .297 .263 .434	.271	.074	.074	.008	1.122	7, 100	.355
2	MusicDuration PROMSmelody MaxPractice listening live digital	-.142 .421 -.529 .094 .175 .209	.561 .329 .051 .317 .249 .415	.268	.072	-.002	.016	1.289	6, 101	.269
3	PROMSmelody MaxPractice listening live digital	.305 -.569 .101 .157 .224	.423 .029 .278 .289 .377	.262	.069	-.003	.023	1.488	5, 102	.2
4	MaxPractice listening live digital	-.525 .092 .164 .220	.039 .318 .267 .386	.250	.063	-.006	.026	1.704	4, 103	.155
5	MaxPractice listening live	-.519 .122 .167	.041 .150 .257	.236	.056	-.007	.028	2.024	3, 104	.115
6	MaxPractice listening	-.415 .135	.079 .109	.209	.044	-.012	.025	2.38	2, 105	.098
7	MaxPractice	-.334	.149	.140	.020	-.024	.010	2.112	1, 106	.149

Note. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B21

Backward multiple linear regression for trombone Vib 1 difference score

Model	Predictors	β	β *sig.*	*R*	R^2	R^2 *change*	*Adjusted* R^2	*F*	*df*	*p*
1	MusicDuration PROMSmelody PROMStuning MaxPractice listening live digital	-.710 -.126 -.072 .305 -.035 .220 -.001	.037 .843 .930 .420 .788 .290 .999	.260	.068	.068	.002	1.027	7, 100	.417
2	MusicDuration PROMSmelody PROMStuning MaxPractice listening live	-.246 .859 -.405 -.178 -.082 .652	.035 .842 .929 .417 .771 .287	.260	.068	.000	.012	1.21	6, 101	.307
3	MusicDuration PROMSmelody MaxPractice listening live	-.714 -.147 .297 -.036 .221	.032 .801 .413 .761 .282	.260	.068	.000	.021	1.465	5, 102	.208
4	MusicDuration MaxPractice listening live	-.752 .291 -.035 .223	.011 .419 .768 .274	.259	.067	-.001	.030	1.832	4, 103	.128
5	MusicDuration MaxPractice live	-.734 .269 .212	.010 .443 .288	.257	.066	-.001	.039	2.435	3, 104	.069
6	MusicDuration live	-.670 .262	.014 .167	.247	.061	-.005	.043	3.369	2, 105	.038*
7	MusicDuration	-.560	.031	.208	.043	-.018	.034	4.757	1, 106	.031*

Note. *$p < .05$. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B22

Backward multiple linear regression for trombone Vib 2 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 change	Adjusted R^2	*F*	*df*	*p*
1	MusicDuration PROMSmelody PROMStuning MaxPractice listening live digital	.025 -.805 -.462 .546 .035 .033 .147	.937 .174 .540 .123 .771 .863 .650	.259	.067	.067	.001	1.014	7, 100	.426
2	PROMSmelody PROMStuning MaxPractice listening live digital	-.787 -.455 .552 .033 .036 .145	.149 .541 .108 .777 .846 .653	.258	.067	.000	.011	1.193	6, 101	.316
3	PROMSmelody PROMStuning MaxPractice listening digital	-.780 -.459 .574 .036 .146	.150 .535 .074 .755 .648	.258	.066	.000	.020	1.438	5, 102	.217
4	PROMSmelody PROMStuning MaxPractice digital	-.805 -.435 .591 .185	.131 .553 .062 .528	.256	.066	-.001	.029	1.789	4, 103	.137
5	PROMSmelody MaxPractice digital	-.948 .532 .192	.046 .075 .511	.250	.062	-.003	.035	2.282	3, 104	.084
6	PROMSmelody MaxPractice	-.969 .557	.041 .060	.242	.058	-.004	.040	3.223	2, 105	.044*

Note. *$p < .05$. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

Table B23

Backward multiple linear regression for trombone Vib 3 difference score

Model	Predictors	*β*	*β sig.*	*R*	R^2	R^2 change	Adjusted R^2	*F*	*df*	*p*
1	MusicDuration	.131	.644	.386	.149	.149	.089	2.476	7, 100	.022*
	PROMSmelody	-1.628	.003							
	PROMStuning	.600	.382							
	MaxPractice	.104	.745							
	listening	.152	.166							
	live	-.308	.080							
	digital	.087	.767							
2	MusicDuration	.123	.661	.385	.148	-.001	.097	2.9	6, 101	.012*
	PROMSmelody	-1.620	.003							
	PROMStuning	.587	.389							
	MaxPractice	.110	.729							
	listening	.163	.107							
	live	-.306	.081							
3	MusicDuration	.144	.598	.384	.147	-.001	.105	3.487	5, 102	.006**
	PROMSmelody	-1.626	.003							
	PROMStuning	.645	.327							
	listening	.169	.089							
	live	-.289	.084							
4	PROMSmelody	-1.520	.002	.381	.145	-.002	.111	4.319	4, 103	.003**
	PROMStuning	.713	.268							
	listening	.161	.099							
	live	-.264	.099							
5	PROMSmelody	-1.250	.004	.367	.134	-.010	.109	5.332	3, 104	.002**
	listening	.176	.071							
	live	-.252	.114							
6	PROMSmelody	-1.347	.002	.336	.113	-.021	.096	6.63	2, 105	.002**
	listening	.142	.136							
7	PROMSmelody	-1.395	.001	.306	.094	-.019	.085	10.869	1, 106	.001*

Note. *p < .05. **p < .01. MusicDuration = length of musical study in years. PROMSmelody = PROMS melody score. PROMStuning = PROMS tuning score. MaxPractice = music practice (hrs/week). Listening = music listening time (hrs/week). Live = music concert attendance per year. Digital = digital music watching (hrs/week).

APPENDIX C

Copyright Permissions

License Number	4216140282081
License date	Oct 25, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cognitive Brain Research
Licensed Content Title	On the neuronal basis for multisensory convergence: a brief overview
Licensed Content Author	M.Alex Meredith
Licensed Content Date	Jun 1, 2002
Licensed Content Volume	14
Licensed Content Issue	1
Licensed Content Pages	10
Start Page	31
End Page	40
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 2, Figure 4
Title of your thesis/dissertation	MUSIC TO OUR EYES: ASSESSING THE ROLE OF EXPERIENCE FOR MULTISENSORY INTEGRATION IN MUSIC PERCEPTION
Expected completion date	Dec 2017
Estimated size (number of pages)	120
Requestor Location	Robert Graham 2 Verano Pl SANTA FE, NM 87508 United States Attn: Robert Graham
Publisher Tax ID	98-0397604

VITA

Graduate School
Southern Illinois University

Robert E. Graham

regraham85@gmail.com

Indiana University Bloomington
Bachelor of Science in Music and an Outside Field, May 2007

Queen's University Belfast
Master of Arts with distinction, Cognition and Culture, December 2010

Southern Illinois University Carbondale
Master of Arts, Psychology, August 2014

Special Honors and Awards:

2017	Arthur R. Menendez Memorial Fund for Graduate Student Vision Research
2016-2017	Dissertation Research Award: Southern Illinois University Carbondale
2016	Runner-up: Paper presentation, Graduate Creative Activities & Research Forum, Southern Illinois University Carbondale
2010	M.A. with Distinction, Queen's University Belfast
2003-2007	Faculty-Awarded Merit Scholarship, Indiana University Bloomington

Dissertation Title:
MUSIC TO OUR EYES: ASSESSING THE ROLE OF EXPERIENCE FOR MULTISENSORY INTEGRATION IN MUSIC PERCEPTION

Major Professor: Dr. Usha Lakshmanan

Publications:

Graham, R. E., Lakshmanan, U. (in press). Tunes & tones: Music, language, and inhibitory control. *Journal of Cognition & Culture, 18(1-2).*

Lakshmanan, U., Graham, R. E. (2016). On the generalizability of chunk-and-pass processing: Perspectives from language acquisition and music. [Peer commentary on "The Now-or-Never bottleneck: a fundamental constraint on language," by M. Christiansen & N. Chater]. *Behavioral and Brain Sciences, 39.*