Music expert-novice differences in speech perception

by

Juan Sebastian Vassallo
BMus., National University of Córdoba (Argentina), 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in Interdisciplinary Studies in the departments of Music and Psychology

University of Victoria

Supervisory committee

Music expert-novice differences in speech perception

by

Juan Sebastian Vassallo
BMus, National University of Córdoba (Argentina), 2013

Supervisory committee

Dr. W. Andrew Schloss (School of Music, University of Victoria)
Supervisor
Dr. James Tanaka (Dept. of Psychology, University of Victoria)
Co-supervisor

Abstract

Supervisory committee

Dr. W. Andrew Schloss (School of Music, University of Victoria)
Supervisor
Dr. James Tanaka (Department of Pyschology, University of Victoria)
Co-supervisor

It has been demonstrated that early, formal and extensive musical training induces changes both at the structural and functional levels in the brain. Previous evidence suggests that musicians are particularly skilled in auditory analysis tasks. In this study, I aimed to find evidence that musical training affects the perception of acoustic cues in audiovisual speech processing for Native-English speakers. Using the McGurk paradigm –an experimental procedure based on the perceptual illusion that occurs when an auditory speech message is paired to incongruent visual facial gestures, participants were required to identify the auditory component from an audiovisual speech presentation in four conditions: (1) Congruent auditory and visual modalities, (2) incongruent, (3) auditory only, and (4) visual only. Our data showed no significant differences in accuracy between groups differentiated by musical training. These findings have significant theoretical implications suggesting that auditory cues for speech and music are processed by separable cognitive domains and that musical training might not have a positive effect in speech perception.

Keywords: *Musical training, Audiovisual speech, McGurk effect*

Table of contents

List of figures

List of tables

Acknowledgments

I would like to gratefully acknowledge the mentorship of my supervisors, Dr. James Tanaka and Dr. Andrew Schloss, for generously sharing their knowledge with me, and for the permanent good energy and good personal treatment. Your invaluable contribution has left a positive mark on my personal and academic growth, and it has been definitely a pleasure to work and learn from you.

Thanks to professors David Clenman, Michael Masson, and Kirk McNally for their contributions to improve this work. Special thanks to Dr. Adam Con, Elissa Poole and Benjamin Butterfield, who kindly invited their students to take part of my study. Thanks to the Director of the School of Music of the University of Victoria, Prof. Christopher Butterfield for giving permission for recruiting music students, and thanks to the Secretary, Alexis Ramsdale, for passing the emails. This research project would not have been possible without their collaboration, and from all the music students who disinterestedly took part of it.

Thanks to my colleagues from the Different Minds Lab, specially to Alison Campbell for her operational collaboration and for her great contributions from the very beginning until the very end of this project. Thanks to Marie Söntgerath and Pascalle Ricard for taking the time to read my work and for giving me valuable feedback. Thanks to Danesh Shahnazian and Patricio Carnelli for their great help with the statistical analysis, and special thanks to Morgan Teskey, Soley Pitre, Michael Chin and Michael Willden for allowing me to use their faces and voices as my stimuli.

Thank you to Doreen Dufresne for being my family in Victoria.

Finally, I want to thank Dr. Daniel Peter Biro, for encouraging me to come to Victoria, for his invaluable collaboration in the first stage of my coursework, and for his willingness for sharing knowledge and for his always smart advice.

Thank you all!

Dedication

A la familia que me dio la vida: Mis viejos y mis hermanas. A los que se fueron, pero están siempre. A los más nuevitos, Tita y Benja.

To the family that life gave me: My parents and my sisters. To those who left but are always present. To the newer ones, Tita and Benja.

Introduction

In the past decades, much research has been conducted on the relationship between musical practice and cognitive abilities, showing that musicians compared to non-musicians demonstrate enhanced performance on several measurements. In this work we are interested in investigating musical training not as a catalyst of cognitive improvements, but as feasible of having a positive effect at the level of auditory perception, particularly in relation to speech, with the aim of contributing with a growing body of research exploring the impact of musical training on language skills. In this introductory chapter I will discuss the overall structure of this thesis, as well as a brief description of each chapter, in order to have a better understanding of the research question and the methodology for carrying out this work.

As a departure point for this thesis I consider important to pose two questions: (1) whether speech and music are related in terms of physical features and perceptual mechanisms, and (2) why one would expect that musical training should result in a perceptual enhancement for speech processing.

Regarding to the first inquiry, the first obvious commonality between *speech* and *music* has to do with their physical medium of propagation: *sound*. Davies (2014) argues that, despite the great diversity and heterogeneity that characterizes music, a universal characteristic emerges from the discussion, and that is the perception of music is subject to more general auditory processing regimes, and that any proposed definition for music should consider it as phenomenon whose medium is *sound organized in time.* The same situation occurs for *speech.* According to Weeks (2019), *speech* is considered just one expression of *language*, specifically related to the physical production and perception of sounds. However, it is known with certainty

that the auditory perception of speech can use the visual information provided by the speaking face, so that it favors the identification and recognition of the message, especially in adverse conditions for hearing, for example, when there is noise in the environment. That visual information, by itself, forms the basis of the code for *lip reading:* a technique of understanding speech by visually interpreting the movements of the speaking face -such as lips, face and tongue, when normal sound is not available. On the basis of these considerations, it can be argued that the main parallelism between *speech* and *music* is the fact that they are both related to sound production and auditory perception.

Even though speech and music are both acoustic and auditory phenomena, Patel (2008) has argued that they are based on different sound systems. In order to better understand the possible differences and similarities between these two, in Chapter 1 I discuss the compounding sound elements of music and speech in terms of their physical properties and perceptual implications. In Chapter 2, I discuss some relevant aspects related to speech production and perception, in order to understand how speech is produced within an anatomical and acoustic framework and how our brain makes use of the available perceptual information to process, segment and extract meaning from the acoustic signal of speech. This discussion is relevant in order to better understand the nature of the proposed study for this research work.

For my experimental procedure, I used an audiovisual speech recognition paradigm based on the phenomenon known as the McGurk effect: A perceptual illusion that occurs when an observer is presented with mismatching speech visual and auditory information that leads to an auditory misperception, e.g. when the auditory component of the syllable /ba/ is paired with visual gestures for the syllable /ga/, a perceiver will probably be led to the perception of an

auditory /da/ (McGurk and McDonalds, 1976). The influence of cues involving facial information perceived by sight, the audiovisual perceptual illusion known as the McGurk effect -the core of our experimental paradigm, as well as the hypothesis of weighted modalities in speech perception are explained in depth in Chapter 3.

In order to tackle my second opening inquiry, whether one would expect that musical training should derive in a perceptual enhancement for speech processing, in Chapter 4, I will carry out a literature review on musical perceptual expertise. First, I will discuss research proposing evidence for auditory enhancements elicited by musical training. Later, I propose, on one hand, evidence supporting the claim that musical auditory training can have a positive effect in the encoding and recognition of acoustic cues relevant for speech perception and intelligibility. On the other hand, I will show evidence disputing the former, by discussing other studies that suggest the feasibility of cognitive and neural dissociations between musical and linguistic processing modules, leading to the assumption that speech and music are not directly linked and one processing system does not affect the other directly. At first sight, these two claims appear to be incompatible, and I consider that my research project aims to shed some light on this particular inquiry.

In my current experiment, I tested participants differentiated by musical training -a sample of *music experts*, consisting of a group of music students of the School of Music of the University of Victoria, and a *control* group composed by individuals that reported no formal musical training during their lifetime. I aimed to find significant group differences in accuracy identifying the auditory component present in an incongruent audiovisual combination based on the McGurk effect. The complete description of the nature and design of my study is

described in chapter 5, as well as the analysis of the experimental data collected, and the discussion of the possible interpretation for the results obtained. A previous study carried out by Proverbio et al. (2016) with the same inquiry is also discussed, as well as *how* and *why* our study aims to tackle some methodological problems present in the former.

I consider important to mention that my personal academic background is in the discipline of music, mainly instrumental performance, music composition and more recently multimedia art creation, and I have found in the exploration of the phenomenon of human perception, a point of origin for a personal creative process. The single idea that information from one sensory modality may alter the perceptual array in another modality, e.g., incongruent visual and auditory information present in the McGurk paradigm, has triggered an artistic inquiry that had been around me since I started working in artistic projects involving technology aimed to generate interactions between sound, visuals and movement. This genuine interest has propelled me to carry out this scientific research, in order to have a deeper comprehension of the phenomenon of *human audiovisual perception*, with the aim of generating new questions regarding to my own artistic practice.

Finally, the main motivation for this work has been to find more evidence that supports the assumption that auditory training developed by musical practice has positive cognitive effects, and thus, further support the use of music education and therapy especially at the level of early schooling due to its cognitive strengthening. I hope that this work can successfully contribute to the universal respect of music as an important discipline in terms of human development.

Chapter 1 – Sound elements for speech and music

Patel (2008) suggests that human beings are exposed since they are born, to a world composed of two distinct sound systems. One of them is *linguistic* and comprises all of the sound categories related to their native languages, such as vowels and consonants. The other one is *musical*, and includes the elements of their musical cultures, such as pitches, timbres and rhythms. He also suggests that music and speech have a very obvious difference in their sound category systems: Whereas *pitch* is the primary basis for sound categories in music (present mostly in intervals and chords), the most salient feature for categories of speech is *timbre* (e.g. vowels and consonants are differentiated by their spectral structure). In this chapter, I will discuss the basic concepts of pitch and timbre, as sound elements present in both domains, speech and music.

Musical pitch

Pitch is defined as the property of a sound that characterizes its highness or lowness to an observer. It is related to, but not identical with, frequency (Law & Rennie, 2015). The measurement of frequency is in cycles per second, or Hertz (Hz). Previous research has shown that the correlation between pitch and frequency is not exact, and that judgments of pitch are affected by other variables such as dynamic, timbre, and duration (Stevens & Volkmann, 1940). According to these authors, pitch is considered a psychological aspect of sound, "one of the dimensions in terms of which we are able to distinguish and classify auditory sensations (…) Pitch differs from frequency in that pitch is determined by the direct response of a human listener to a sound stimulus, whereas frequency is measured with the help of measure

instruments" (p 329). According to Patel (2008), "pitch is the most common dimension for creating an organized system of musical elements" (p. 13). He also argues that pitches are present in all cultures around the world, similarly organized in the form of groups arranged in some stable and discrete order. This set of distinctive pitches are known as *musical scales*, and serve as reference point in the creation of musical patterns. A musical scale consists of a particular choice of pitches separated from each other by certain intervals within some range - usually an *octave.* In acoustic terms, an *octave* is the interval between two musical pitches that have frequencies in the ratio 2:1: A pitch with twice or half the frequency of another pitch, with which it usually shares a common denomination or name (e.g. a = 220, a'= 440, a'' = 880). Individual pitches can be combined simultaneously to create new sonic entities, such as intervals and chords, that have distinctive perceptual qualities.

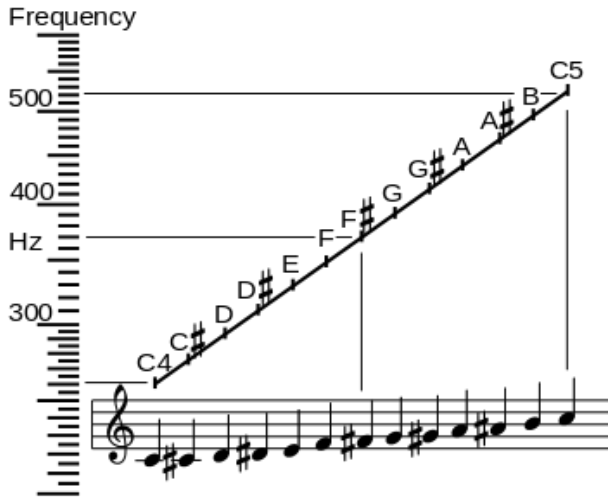


Figure 1 – Pitches noted on a staff and their corresponding frequency on a logarithmic scale.

Patel (2008) discusses the fact that there is a great cultural diversity in musical scale systems, and classifies this diversity in four types: (1) Amount of "tonal material" (p. 17) within each octave for choosing pitches. As an example of this, the author suggests a comparison

between the division in 12 equally-distanced pitches present in Western music and the division in 22 unequal parts present in Indian classical music; (2) Number of pitches chosen by octave: Although it can vary widely, a common range across cultures has been observed to be between 5 and 7. (3) different interval patterns between pitches: the author suggests that the interval usually ranges from 1 to 3 semitones[1], but an interval of 2 semitones (1 tone) has been observed as common standpoint, and an asymmetric distribution of these intervals along the octave is considered a commonality present in most cultures; and (4) different tuning systems: A *tuning system* is used to define *which* pitches to use, by the choice of number and spacing of frequency values used for each pitch in that system. One of the most salient features of the tuning system in Western music is that it is based on a fixed reference, usually 440 Hz for the pitch A, and the rest of the pitches are also based on fixed frequencies based on an intervallic relation to 440 Hz.

Pitch contrasts in speech

According to Patel (2008), the most relevant physical correlation between pitch and speech is the *fundamental frequency* of vocal fold vibration (F0). This attribute is known to convey linguistic information, especially in the denominated *tone languages*, but it is considered mostly an attitudinal and emotional marker for speech. Patel (2008) defines a tone language as "a language in which pitch is as much a part of a word's identity as are the vowels and consonants, so that changing the pitch can completely change the meaning of the word" (p. 40). Although most of Western languages are non-tonal, Fromkin (2014) has argued that over

[1] A semitone is the smallest musical interval commonly used in Western tonal music.

half of the world's languages are tonal, including the majority of languages in Africa and southeast Asia. As my research work is focused on English language –non-tonal, I will discuss the use of pitch contrast in languages that are non-tonal, thus, an in-depth discussion of the use of pitch in tonal languages is beyond the scope of this work.

The most relevant linguistic information conveyed by the use of pitch contrasts in speech is related to *intonation*, a linguistic function that is used for indicating the attitudes and emotions of the speaker, and for signaling the difference between statements and questions, and between different types of questions. Wells (2006) has proposed a list of 6 distinct functions of intonation for English language: (1) Attitudinal function: For expressing emotions and attitudes; e.g.: A fall from a high pitch on the 'mor' syllable of "good morning" suggests more excitement than a fall from a low pitch; (2) grammatical function: To identify grammatical structure. e.g.: It is claimed that in English a falling pitch movement is associated with statements, but a rising pitch turns a statement into a yes/no question, as in *He's going ↗**home**?* (3) Focusing: To show what information in the utterance is new and what is already known. E.g.: in English *I saw a ↘**man** in the garden* answers "Whom did you see?" or "What happened?", while *I ↘**saw** a man in the garden* answers "Did you hear a man in the garden?"; (4) discourse function: To show how clauses and sentences go together in spoken discourse. E.g.: Subordinate clauses often have lower pitch, faster tempo and narrower pitch range than their main clause, as in the case of the material in parentheses in "The Red Planet (as it's known) is fourth from the sun" (5) Psychological function: To organize speech into units that are easy to perceive, memorize and perform. E.g.: the utterance "You can have it in red blue green yellow or ↘black" is more difficult to understand and remember than the same utterance divided into tone units as in "You

can have it in ↗red | ↗blue | ↗green | ↗yellow | or ↘black"; and (6) indexical function: To act as a marker of personal or social identity. E.g.: Group membership can be indicated by the use of intonation patterns adopted specifically by that group, such as street vendors or preachers.

Pierrehumbert (1987) developed a system for analysis of intonation, widely known as ToBI (short for "Tones and Break Indices"). The most important point of this system is that only two tones, associated with pitch accents, are recognised, these being H (high) and L (low), and indicate relative highs and lows in the intonation contour. All other tonal contours are made up of combinations of H, L and some other modifying elements. Ladd (2001) suggests that speech tones are scaled relative to two reference frequencies corresponding to the top and bottom of an individual's speaking range. This range can vary between speakers, and can be elastic within a speaker, for example, growing when speaking loudly or with strong positive affect. Ladd argues that what stays relatively constant across contexts and speakers is pitch level as a proportion of the current range. Furthermore, Ladd and Morton (1997) discuss that there is a categorical difference between *normal* and *emphatic* pitch accent peaks in English, rather than a continuum of gradually increasing emphasis.

Musical timbre

According to Wallmark (2014), *timbre* is defined as "the attribute of musical sound that distinguishes one source from another when pitch and loudness are held constant. Also known as *tone*, *tone quality*, or *tone color*." (p. 2). This distinctiveness of a sound is usually the result of the presence of a structure of *overtones*[2] that altogether compound a more or less complex

[2] The term *overtone* is used to refer to any resonant frequency above a fundamental frequency.

waveform, usually different and unique for each source in a particular situation. For example, a note from a musical instrument will have several *harmonics*[3] present, depending on the type of instrument and the way in which it is played. In Figure 2, the peaks on the frequencies that integer multiples from the fundamental can be clearly observed (harmonics), together some other random frequencies at a minor amplitude.

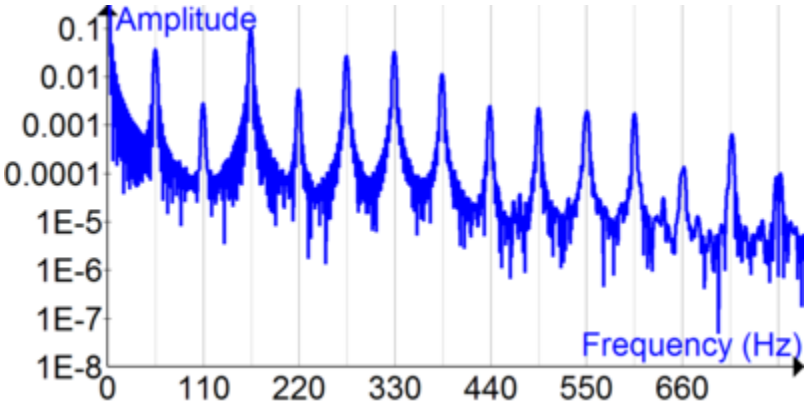


Figure 2 – Spectral analysis of the sound of a bass guitar from an open string A (55 Hz).

Sounds are usually categorized according to their timbre, in *harmonic* and *inharmonic* varieties. Those with harmonic spectra contain most of their energy in the harmonic series (integer multiples of the fundamental frequency, e.g., 440 Hz, 880 Hz, or 1,320 Hz), while the overtones of inharmonic spectra are less regularly dispersed across the spectrum. The great majority of musical instrumental sounds are largely harmonic. Wallmark (2014) argues that the definition of timbre as dependent on the component structure of overtones or *spectrum*[4] of a

[3] A *harmonic* is an integer (whole number) multiple of the fundamental frequency of a vibrating object. The term *harmonic* is contained within the definition of *overtone*, in the sense that an overtone may or may not be a *harmonic*.

[4] The distribution of pure tones in a sound wave, usually represented as a plot of power or amplitude as a function of frequency.

sound has expanded to include a range of other acoustic attributes, and in addition to spectral characteristics, timbre is known to depend on temporal properties such as *dynamic envelope*[5], as well as *transients*[6] specific to each sound producer (e.g., the noise of the violin bow or the "blatty" attack of the trombone).

Patel argues that timbre is rarely used as the basis for organized sound contrasts in music, mainly for two reasons. The first one is that great changes in the timbre of an instrument usually require also great changes in the way it is excited, and for some instruments this is difficult or literally impossible without altering its physical properties. A second reason is that timbral contrasts are not organized in a system of orderly perceptual distances from one another, for example, in terms of "timbre intervals". The author suggests that these perceptual distances present in pitch, characterized as intervallic relations between notes "allow higher-level relations to emerge" (p. 33), whereas in timbre, those type of relations are perceptually more difficult to establish. However, some musical expressions have emerged that are known to be based on timbral contrasts. One of the most well-known examples in Western music is Arnold Schoenberg's *Klangfarbenmelodie*: A musical composition system in which the series of musical notes are replaced by specific timbral values, so that instead of successive note changes, the instrument is changed –mostly over a same fixed note, or a single melodic pattern. In this system, instruments are used only depending on their timbre. This compositional system was used by the end of the 19th century and the beginning of the 20th century, especially in works by Arnold Schönberg and Anton Webern. Later on, the development of electronic music and

[5] The dynamic envelope of a sound accounts for changes in the energy of its acoustic wave perceived by humans over time. It is usually stated that it is divided in four parts: attack, decay, steady-state, and release.

[6] Transients are defined as any rapid change in sound level.

the creation of synthetic sound, allowed composers to experiment with new compositional systems based on Schoenberg's idea, but without the constrain of the physical limitation of acoustic instruments. Still, musical systems based on timbre are far from being as used as those based on pitch.

Timbre contrasts in speech

Patel (2008) has suggested that "speech is fundamentally a system of organized timbral contrasts", and that "timbre is the primary basis for linguistic sound categories" (p. 51). The human voice is a great source for generating timbral contrasts, as these contrasts result from continuous changes in the vocal tract as speech sounds are produced. Within the next chapter, I will discuss the basis for mechanisms of speech production, in order to understand how speech is produced within an anatomical and acoustic framework.

In summary, salient differences between music and speech are that the amount of change in spectral shape, as each syllable contains rapid changes in overall spectral shape which help cue the identity of its constituent phonemes, whereas timbral changes within musical notes occur to a lesser extent than within spoken syllables. Differences between instrumental music and speech also extend to patterns of fundamental frequency (F0), with the most salient difference the lack of stable pitch and intervals in speech.

Chapter 2 – Fundamentals of speech production and perception

Speech production

The respiratory system together with the vocal organs are known to be a set of important physiological structures in the production of speech sounds, particularly the *vocal folds*. Redford (2015) describes the vocal folds as "soft tissue structures contained within the cartilaginous framework of the larynx, and serve as the primary generator of sound for vowels as well as a pressure controller for many consonants" (p. 54).

In order to start the speech's acoustic signal production sequence, an airstream from the lungs passes between the vocal folds. In the case that vocal cords are apart (Fig. 1 a), as they normally are when breathing out, the air from the lungs passes relatively free by the pharynx and the mouth. But if the vocal cords are close together (Fig. 1 b) so that there is a narrow passage between them, the airstream will cause a pressure below them that will build up until they are blown apart again. The flow of air between them will then cause them to be sucked together again, creating a vibratory cycle.

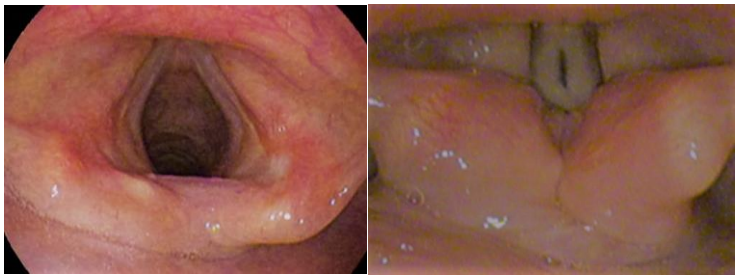


Figure 3 – (a) Vocal folds (open) (b) Vocal folds (phonating).

A particular configuration of the vocal tract (the larynx and the pharyngeal, oral, and nasal cavities) resulting from the positioning of the mobile organs (e.g., tongue) relative to other parts

of the vocal tract that may be rigid (e.g., hard palate) is known as *articulation,* and each specific configuration modifies an airstream to produce different speech sounds. The main articulators are the *tongue*, the *upper lip*, the *lower lip*, the *upper teeth*, the *upper gum ridge* (alveolar ridge), the *hard palate*, the *velum* (soft palate), the *uvula* (free-hanging end of the soft palate), the *pharyngeal wall*, and the *glottis* (space between the vocal cords).

Articulations may be divided into two main types: (1) ***Primary articulation*** refers to either (a) the place and manner in which the stricture is made for a *consonant* or (b) the tongue contour, lip shape, and height of the larynx used to produce a *vowel*. The primary articulation may still permit some range of movement for other articulators not involved in its formation; and (2) S***econdary articulation***, a type of articulation that involves freedom in one of the articulators (e.g., an "apico alveolar" articulation involves the tip of the tongue but leaves the lips and back of the tongue free to produce some degree of further stricture in the vocal tract). Among the most used secondary articulations are ***palatalization*** (the front of the tongue approaching the hard palate); ***velarization*** (the back of the tongue approaching the soft palate, or velum); ***labialization*** (added lip-rounding), ***glottalization*** (complete or partial closure of the vocal cords); and ***nasalization*** (simultaneous passage of air through the nasal and oral tracts, achieved by lowering the velum).

Segmentation

Auditory speech perception is dependent upon acoustic information available to the brain. The process of perceiving speech begins at the level of the sound signal and the process of audition. Tatham and Morton (2006) have proposed that *speech* consists of a string of discrete sounds, and that all utterances are to be regarded as linear permutations or rearrangements into

strings of a small number of such sounds. They suggest that the norm for a single language is having around 45–50 of these discrete sounds. These discrete sound units are known as *phonemes*. According to Colman (2015), a phoneme is a speech sound with an average duration of 70 to 80 milliseconds at normal speaking rate, that distinguishes one word from another in a given language. Colman also suggests that a phoneme may have various phonetically distinct articulations, known as *allophones*, that are regarded as functionally identical by native speakers of the language. Phonemes are defined by distinctive features that are relevant or significant as they allow a contrast to be made between phonological units of a given language, especially the phonological attributes that distinguish minimal pairs. Colman (2015) defines a *minimal pair* as pair of words that differ in only one speech sound but have distinct meanings, thus establishing that the speech sounds in question are different phonemes. For example, the fact that *cap* and *cab* have different meanings establishes that /p/ and /b/ are different phonemes in English. In English, /r/ and /l/ give distinct meanings to minimal pairs (such as *row* and *low*) and are therefore distinct phonemes, whereas in Japanese they do not and are therefore allophones, which explains why they are often confused by native Japanese speakers of English, and that therefore underlie the definition of a phoneme.

Greenberg (2006) argues that a higher level in the hierarchy of speech in which phonemes are grouped, is the *syllable*. This author proposes that a syllable consists of an optional *onset* containing between zero and three consonants, an obligatory *nucleus* composed of a vocalic sound, which can be either a monophthong like the *a* in “at”, or a diphthong, like the *ay* in “may”, and an optional *coda*, containing between zero and four consonants. English syllables are typically 100 to 500 ms long, and are characterized by an *arc-shaped* dynamic envelope,

since the vowel nucleus is normally up to 40 dB more intense than the consonants of the onset or the coda (see figure X in which a yellow line shows the arc-shaped dynamic envelope for the syllable /pa/).

Acoustic cues

After processing the auditory signal, speech sounds are further processed to extract phonetic information. The speech sound signal contains a number of acoustic cues that differentiate speech sounds belonging to different phonetic categories. Gussenhoven and Jacobs (2005) observe that phonemes are usually analyzed as bundles of *distinctive features*. One motivation for this level of analysis is that the phonemes of a language have relationships of similarity and difference in terms of the way they are produced.

One of the most studied acoustic cues in speech perception is *voice onset time* (VOT), a primary cue that is used by our perception to differentiate between *voiced* and *unvoiced* plosives sounds, such as [b] and [p], and [k] and [g]. The two phonemes [p] and [b] are similar in many respects, both involving a closure of the lips followed by a rapid release and the onset of vocal fold vibration. VOT is defined as the length of time that passes between the release of a stop consonant and the onset of voicing, arising from the vibration of the vocal folds, characterized by some *periodicity* present in the acoustic wave. An acoustic analysis of the sounds [p] and [b] before a following vowel shows that the VOT is much shorter in the latter than in the former. Thus, the two phonemes can be analyzed as sharing a number of articulatory features, and differing in the voicing feature.

Other cues used to differentiate sounds include the acoustic result produced by the articulation of the airflow at different places and in different manners. The different

configurations of the vocal tract that occur while a stream of air passes through give rise to different categories of speech sounds, known as *vowels* and *consonants*.

Consonants

Consonants are speech sounds made by obstructing the glottis (the space between the vocal cords) or oral cavity (the mouth) and either simultaneously or subsequently letting out air from the lungs (McArthur, T., Lam-McArthur, J., & Fontaine, L., 2018). Consonants are discussed in terms of three anatomical and physiological factors: (1) the state of the glottis (whether or not there is *voicing* or vibration in the larynx), (2) the place of articulation (that part of the vocal apparatus with which the sound is most closely associated), (3) and the manner of articulation (how the sound is produced). When the air turbulence generated at an obstruction involves random (aperiodic) pressure fluctuations over a wide range of frequencies, the resulting noise is called *aspiration* if the constriction is located at the level of the vocal folds, as for example during the production of the sound [h]. If the constriction is located above the larynx, as for example during the production of sounds such as [s], the resulting noise is called *frication*. The explosion of a plosive release also consists primarily of frication noise.

Vowels

Vowels are speech sounds characterized by absence of obstruction or audible friction in the vocal tract (allowing the breath free passage), and typically composed of periodic or quasi-periodic sound produced by modulation of the airflow from the lungs by the vocal folds (McArthur, T., Lam-McArthur, J., & Fontaine, L., 2018). The quality of a vowel is mainly determined by the position of the tongue and the lips. Each configuration modifies the acoustic excitation signal (airflow) causing some frequencies to resonate and some frequencies to

attenuate. A *formant* is a peak in the spectrum of frequencies of a specific speech sound, analogous to the fundamental frequency or one of the overtones of a musical tone, which helps to give the speech sound its distinctive sound quality or *timbre* (Colman, 2015). Vocalic sounds consist of a fundamental frequency (F0) and its harmonic components (F1, F2, F3, etc.), and the distinctions perceived between vowels lie, in particular, in the *location* on the *frequency range* of the set of formants, particularly of the lower three, and in their amplitude (energy).

Spectrogram

A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time. According to Colman (2015), a speech spectrogram is defined as "a graph of the harmonic spectrum of speech sounds (…) showing sound frequency on the vertical axis and time on the horizontal axis" (p. 15). A third dimension indicating the amplitude (energy) of a particular frequency at a particular time is represented by the intensity or color of each point in the image. Figure 4 shows the spectrogram of the syllable /pa/ pronounced by a female voice, obtained using the software Praat 6.0.42. In this spectrogram, some of the relevant cues for speech recognition can be clearly observed: (1) The noise burst known as Voice Onset Time (VOT), with a duration of approximately 10 milliseconds, indicating the presence of a *voiceless* sound, in this particular case, the consonant [p], characterized as a *plosive* articulated *bilabially*. The offset of the voiceless noise burst and the start of the vowel's loudness peak has been marked with a red dashed line; (2) the formant structure for the vowel [a] has been marked with lines of red dots, representing resonating frequencies along the time axis; (2) the fundamental frequency F0 is marked with a blue line along the time axis, and the dynamic energy arc for the syllable is represented by a yellow arc-shaped line.

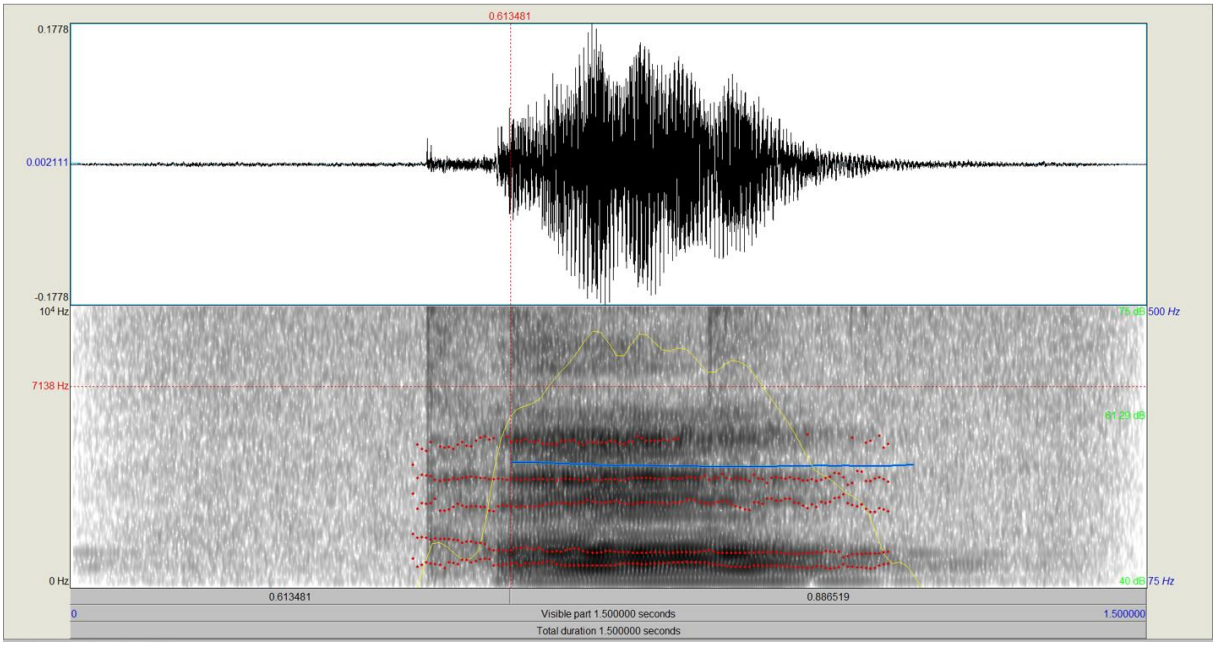


Figure 4 – Amplitude plot (above) and spectrogram (below) of a syllable /pa/ obtained with the software Praat 6.0.42.

The IPA alphabet

Linguists have developed efficient ways to represent speech sounds in writing, using a standardized system known as the *International Phonetic Alphabet*, or IPA (See appendix 1). According to the Handbook of the International Phonetic Association (1999), Its purpose is "to provide, in a regularized, accurate and unique, the representation of the sounds of any oral language, 1 and in the professional field is used by linguists, speech therapists and therapists, foreign language teachers, lexicographers and translators" (p. 10). The symbols of the International Phonetic Alphabet are divided into three categories: (1) *letters, that i*ndicate basic sounds (vowels and consonants); (2) *diacritics*, that specify those sounds; and (3) *suprasegmental*, that indicate qualities such as speed, tone and accentuation. Diacritics and suprasegmentals are divided according to whether they indicate articulation, phonation, tone,

intonation or accentuation. The IPA alphabet can be used to transcribe any language in the world. It is the most widely used phonetic alphabet in the world.

Phonological processing

Massaro (2001) proposes that humans "perceive speech as a discrete auditory message composed of words, phrases, and sentences (…). Somehow, this continuous signal is transformed into more or less a meaningful sequence of discrete events" (p. 14870). In linguistics, this process is known as *phonological processing*, and refers refers to the use of phonological information (i.e., the sounds of one's language) in processing written and oral language (Wagner & Torgesen, 1987).

Wagner & Torgesen (1987) discuss three kinds of phonological processing: (1) *phonological awareness;* (2) *phonological recoding in lexical access;* and (3) *phonetic receding in working memory. Phonological awareness* is the awareness of the sound structure of a language and the ability to consciously analyze and manipulate this structure. This is achieved via a range of tasks, e.g. the capacity for speech sound segmentation, and putting together sounds presented in isolation to form a word. Phonological awareness includes *phonemic awareness,* which applies when the units being manipulated are phonemes, rather than words, onset-rime segments, or syllables. *Phonological recoding in lexical access*, refers to the process of "getting from a written word to its lexical referent by recoding the written symbols into a sound-based representational" (p. 92). This process involves storing phoneme information in a temporary, short-term memory store. An example of a task involving phonological working memory is the repetition of non-words, e.g., repeat /pæg/. Finally,

Phonological retrieval is the ability to recall the phonemes associated with specific *graphemes*[7], which can be assessed by rapid naming tasks (e.g., rapid naming of letters and numbers).

Lack of invariance

Speech is a dynamic phenomenon and that it is unlikely to find constant relations between a phoneme of a language and its acoustic manifestation. This *lack of invariance* is driven mostly by three causes: (1) *phonetic environment*: Speech sounds do not strictly follow one another, rather, they overlap. A speech sound is influenced by the ones that precede and the ones that follow. The phenomenon known as co-articulation (Matthews, 2014) refers to the simultaneous or overlapping articulation of two successive phonological units; and (2) *differing speech rate:* Many phonemic contrasts are constituted by temporal characteristics (short vs. long vowels or consonants, voiced vs. voiceless plosives, etc.) and they are certainly affected by changes in speaking tempo (Nygaard & Pisoni, 1995); and (3) *different speaker identity:* Phonologically identical utterances show a great deal of acoustic variation across speakers, and listeners are able to recognize words spoken by different talkers despite this variation. There is evidence for a *normalization* process that adjusts for variations in the voice quality of different speakers and speaking rates.

Perceptual normalization

During the perceptual normalization process, listeners filter out the inter-source (speaker) and inter-stimulus (sound) variation to arrive at the underlying category. An example of an inter-source variation consists of vocal-tract-size differences, that result in formant-frequency

[7] A grapheme is defined as the smallest unit in the written form of a language, usually a letter or combination of letters representing a single phoneme, such as the *b* in *book*, the *s* in *sip*, the *sh* in *ship*, or the *ph* in *photograph*.

variation across speakers. In order to resolve this situation, a listener is required to adjust his or her perceptual system to the acoustic characteristics of a particular speaker. This process has been called vocal tract normalization, and According to Syrdal and Gopal (1986), this may be accomplished by considering the ratios of formants rather than their absolute values. Similarly, in terms of speech rate, the listener can potentially resolve some of this acoustic variability using rate normalization. According to Jaekel, Newman, & Goupell (2017), rate normalization is the process by which the perception of speech sounds with similar acoustical properties is altered on the basis of sentence context and the speaker's rate of speech.

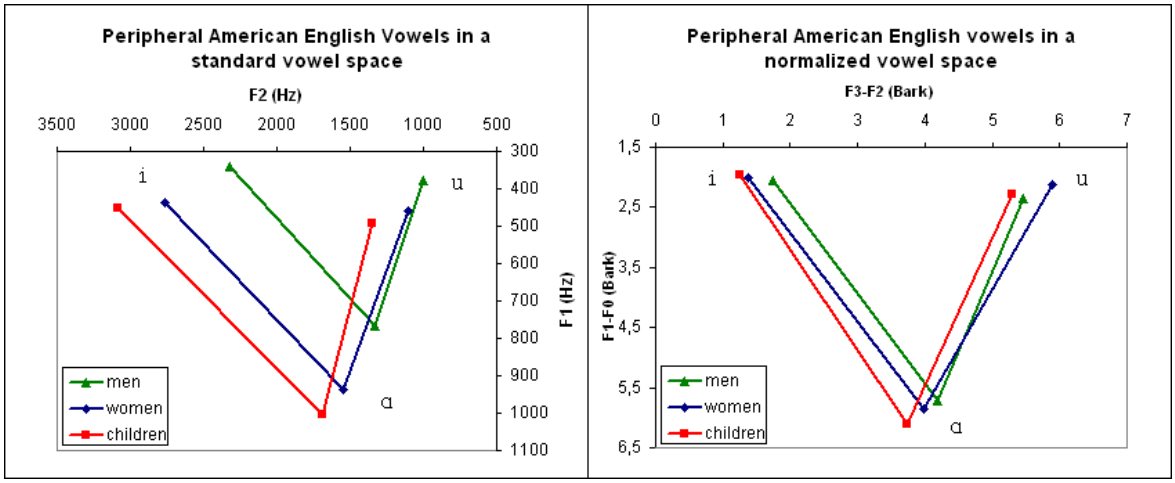


Figure 6 – American English vowels /i/, /a/, and /u/ in a standard F1 by F2 plot (left panel) and in a plot showing formant distances rather than absolute values (right panel).

Oden and Massaro (1978) have proposed a model for perception of speech that provides a detailed description of the processes that may be involved in using featural information to identify speech sounds. Figure 7 presents a schematic diagram of the auditory recognition process in Massaro's model. According to this model, the auditory stimulus is transduced by the auditory receptor system and acoustic features are detected and stored in *preperceptual auditory storage* (PAS). The features stored in PAS are a direct consequence of the properties

of the auditory stimulus and the auditory receptor system. The features are assumed to be independent and the value of one feature does not influence the value of another at this stage of processing. The primary recognition process evaluates each of the acoustic features in PAS and compares or matches these features to those that define perceptual units in *long-term memory* (LTM). Every perceptual unit has a representation in LTM, which is called a *sign* or *prototype.* The prototype of a perceptual unit is specified in terms of the acoustic features that define the ideal acoustic information as it would be represented in PAS. The recognition process operates to find the prototype in LTM that best matches the acoustic features in PAS.

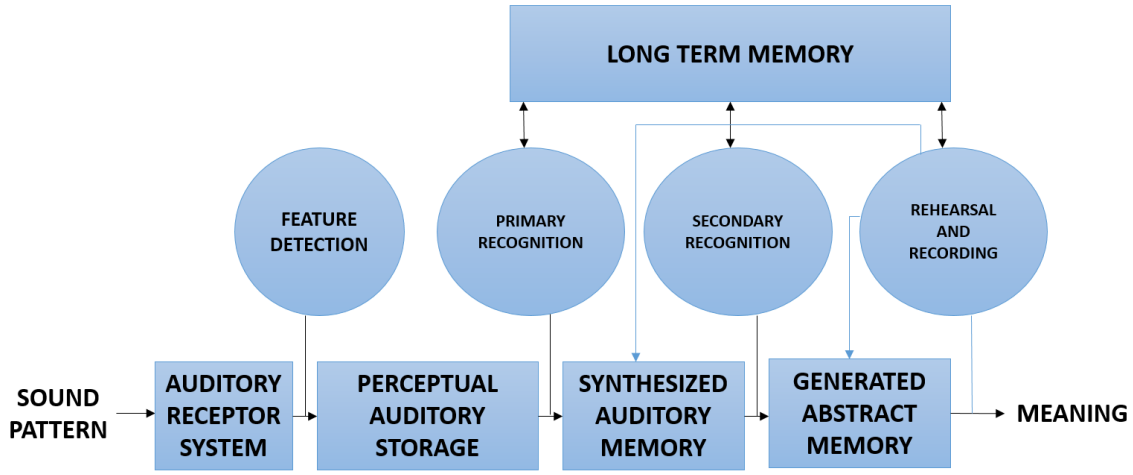


Figure 7 - Schematic diagram of the general auditory information processing model (Oden and Massaro, 1978).

This model has been denominated by Massaro as the *Fuzzy Logical Model of Speech perception*, and it assumes that "(1) acoustic cues are perceived independently, (2) the feature evaluation provides information about the degree to which each quality is present in the speech sound, (3) each speech sound is defined by a propositional prototype in long-term memory that determines how the featural information is integrated, and (4) the speech sound is identified on the basis of the relative degree to which it matches the various alternative prototypes" (Oden

and Massaro, 1978 p. 173). According to the proposed integration model, there are three conceptually distinct operations involved in phoneme identification: (a) The *feature evaluation* operation determines the degree to which each feature is present in PAS, (b) the *prototype matching* operation determines how well each candidate phoneme provides an absolute match to the speech sound, and (c) the *pattern classification* operation determines which phoneme provides the best match to the speech sound relative to the other phonemes under consideration. However, since perception is a noisy process in which a given physical stimulus will be perceived differently at different times, phoneme classification is necessarily a probabilistic process. An extended explanation of the Fuzzy Logical Model of Speech perception, also involving visual cues, as well as a proposed probabilistic model to be used for classifying speech stimuli is described in next chapter.

Chapter 3 - Audiovisual integration in speech perception

Even though the perception of language is primarily dominated by audition, it is known that humans use visually mediated information to facilitate communication, especially in noisy conditions. According to Schroeder (2008), visual speech perception is still viewed primarily as an influence on auditory speech perception, and the literature is consistent with the view that visual speech stimuli are phonetically impoverished, but that the phonetic information is not so reduced that accurate visual spoken word recognition is impossible. A spoken word can be recognized despite phonetic impoverishment, if it is sufficiently distinct from other words in the mental lexicon, and visual phonetic information can be sufficiently distinct. Researchers had reported that under noisy conditions, enhancements to auditory speech intelligibility and language comprehension occur when the listener can also view the talker (Sumby and Pollack 1954; Sommers & Phelps, 2016).

Visemes

Speech production simultaneously produces the sounds and sights of speech. Both optical and acoustic phonetic attributes instantiate speech features on the basis of diverse sensory information, but the visual information for every phoneme cannot be inferred accurately from a simple one-to-one mapping between the visibility of speech production anatomy (e.g. lips, mouth, tongue, glottis) and acoustic speech features (e.g. voicing, place, manner, nasality). This is because the vocal tract shapes, glottal vibrations, and velar gestures that produce acoustic speech are not all directly visible (Stevens, 1998). The concept of *viseme* (Fischer, 1968) was invented to describe and account for the somewhat stable patterns of lip readers' phoneme

confusions. Visemes are typically formed using some grouping principle such as hierarchical clustering of consonant confusions from phoneme identification paradigm. In the absence of auditory input, lip movements for /ga/ are frequently misread as /da/, while those for /ka/ are sometimes misread as /ta/; /pa/ and /ba/ are often confused with each other, but never misread as /ga/, /ta/, /ka/ or /da/.

Viseme groups	
Group	Cluster
1	/p, b, m/
2	/f, v/
3	/w/
4	/l/
5	/sh, zh/
6	/th, dh/
7	/r/
8	/n, y, d, g, k, s, z, t/

Table 1 – Visemes (Williams, Rutledge, Garstecki, & Katsaggelos, 1997)

The McGurk effect

The so-called McGurk effect (McGurk and McDonald 1976), is an auditory perceptual illusion that occurs when visual speech information conflicts with auditory speech information, provoking the illusion of hearing a different –not present- phoneme. According to Tiippana (2014), the McGurk effect should be defined as a "categorical change in auditory perception induced by incongruent visual speech information, resulting in a single percept of hearing something other than what the voice is saying. When the McGurk effect occurs, the observer has the subjective experience of hearing a certain utterance, even though another utterance is presented acoustically" (p. 1).

There are two typical responses in the McGurk effect. The combination of 'bilabial' (involving both lips) sounds and 'velar' (back of the tongue in contact with the soft palate) mouth movement typically results in a 'fusion response', in which a new phoneme different from the originals is perceived. For instance, when an auditory /ba/ is dubbed with visual /ga/, or auditory /pa/ dubbed with visual /ka/, subjects usually perceive the phoneme as a /da/ or /ta/ respectively, and this is referred as 'fusion' response. On the other hand, when an auditory /ga/ is dubbed with visual /ba/, or auditory /ka/ dubbed with visual /pa/, subjects usually perceive the phoneme as a /bga/ or /pka/, and this is referred as the 'combination' response[8]. The percentages in fig. 8 are those obtained for each response in the original McGurk experiment (1976). It can be observed an asymmetry in terms of illusion sensitivity, being the 'fused' responses more likely to be experienced in comparison with the 'combined' ones.

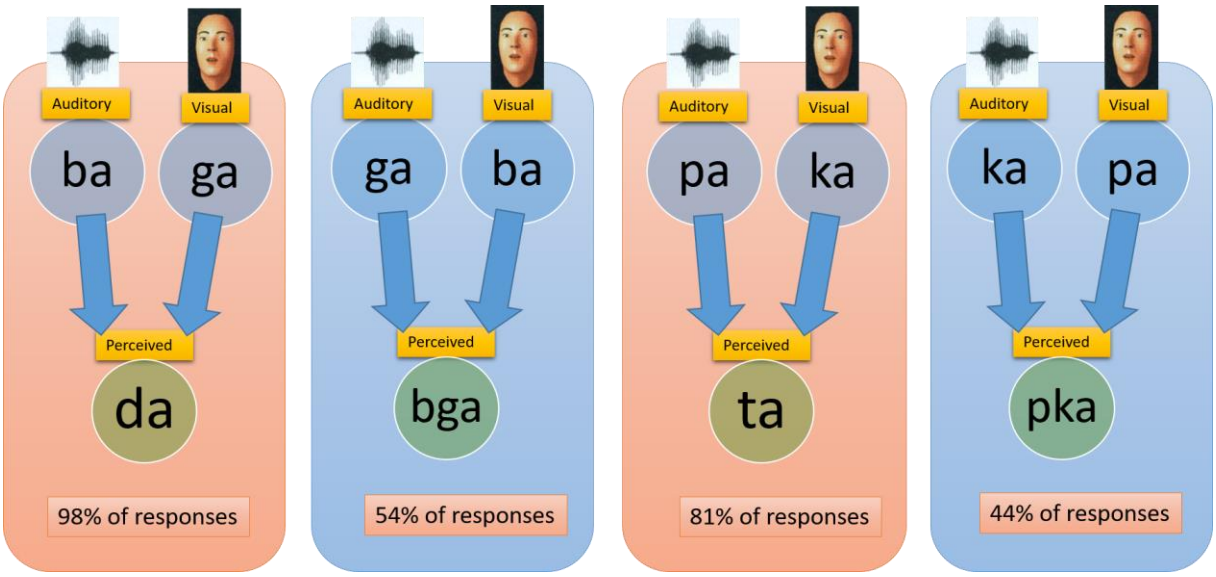


Figure 8 – Reported perceived sounds from the combination of incongruent audio and visual information.

[8] The words 'Fusion' and 'combination' for the different illusory responses present in the McGurk effect were proposed by McGurk and McDonalds (1976).

The appearance of McGurk effect may be related to the language development in general. The original study by McGurk & MacDonald (1976) showed that preschool and school children (i.e. 3–5 and 7–8-year-old groups) reported a weaker McGurk effect than did the adults. Several studies (Sekiyama & Tohkura 1991; 1993; Sekiyama 1997;) have also proposed evidence for differential occurrences of the McGurk effect across various languages (e.g. Japanese, American and Chinese). Sekiyama (1997) has suggested that auditory cues, rather than visual cues, are more powerful in recognizing lexical tones in Chinese. Consequently, Chinese speakers might rely more on auditory information when perceiving inconsistent stimuli and manifest weaker McGurk effects than their Japanese counterparts. Hisanaga et. al (2016) suggested that native English speakers are influenced by visual mouth movement to a greater degree than native Japanese speakers when listening to speech. These results clearly indicate the impact of language and/or culture on multi-sensory speech processing, suggesting that linguistic/cultural experiences lead to the development of unique neural systems for audiovisual speech perception, although some recent research has disputed these findings, showing evidence supporting the claim that high individual variability in perception of the McGurk effect necessitates the use of large sample sizes to accurately estimate group differences (Magnotti et al., 2015), and by using population samples of ~300 individuals, cross-language differences tend to be not significant.

The Fuzzy Logic Model of Perception

One of the most relevant models that attempt to explain how speech perception works is the Fuzzy Logic Perceptual Model of speech (FLMP), developed by D. Massaro (1978, 1980, 1987). According to this model, speech perception can be understood as a problem of

classifying the features that are present in a perceptual pattern. Massaro proposes that the concepts that people use to categorize all sorts of objects have fuzzy boundaries: "A category has fuzzy boundaries in the sense that we consider some things to be strong members of the category, some weak members and others as nonmembers" (p. 188). In a newer version of the FLMP discussed in Chapter 2, Massaro (1987) expands the original model by adding visual cues to the acoustic ones. Thus, in order to recognize a speech stimulus, categorization process takes place in three stages: (1) **Evaluation**: The features of the acoustic signal are analyzed, and in parallel, visual information compounded of vocal and facial motion is evaluated; (2) **Integration:** We match the features of a given speech signal with the features of the prototypes that are stored in memory and we attempt to determine which prototype best integrates the presented configuration; and **(3) Decision:** The sound is classified as the pattern that best fits the features of the stimulus that was presented to us. The author also proposes that particularly for speech, human ability to classify patterns is extremely fast.

In figure 9, the three processes are shown proceeding left to right in time to illustrate their necessarily successive but overlapping processing. The sources of information are represented by upper-case letters. Auditory information is represented by A_i and visual information by V_i. The evaluation process transforms these sources of information into psychological values (indicated by lowercase letters a_i and v_i). These sources are then integrated to give an overall degree of support, S_k, for each speech alternative k. The decision operation maps the outputs of integration into some response alternative, R_k. The response can take the form of a discrete decision or a rating of the degree to which the alternative is likely (Massaro, 2001).

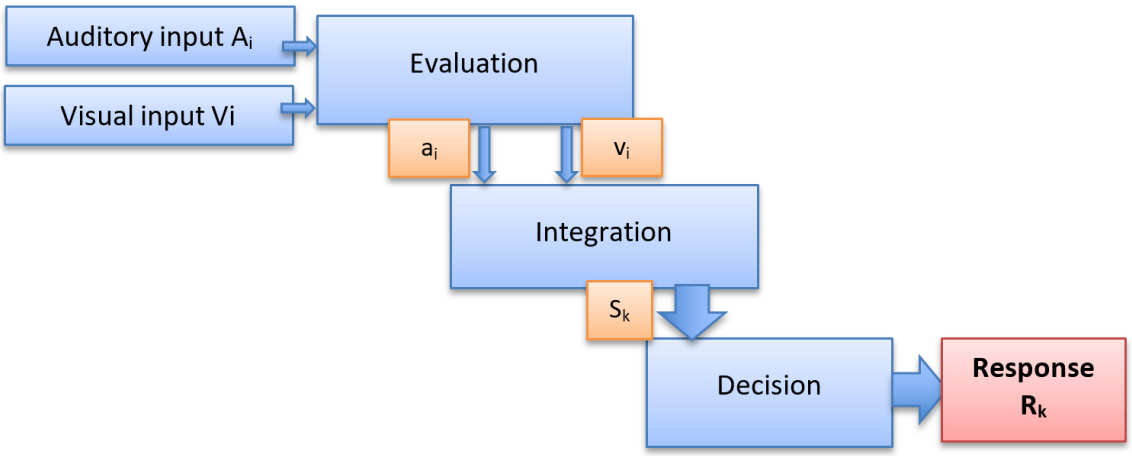


Figure 9 – Schematic representation of the three processes involved in perceptual recognition (Massaro, 2001).

The Fuzzy Logic Model of Perception intends to explain the McGurk effect proposing the idea that perceivers tend to interpret an event in a way that is most consistent with all the sensory information, in audiovisual speech, both sight and sound. Previous research using matrices of confusion[9], has shown that auditory /ba/ is seldom confused with /ga/, and that visual /ga/ looks significantly dissimilar to /ba/, thus unlikely to be confused, but auditory /ba/ shows a significant extent of confusability with /da/, and visual /ga/ is also confusable with /da/, so according to this principle, the most suitable response for an audiovisual configuration such as the classical mismatching McGurk [Auditory: /ba/ Visual: /ga/] would be /da/. The matrix of confusion for the congruent bimodal presentation shows that syllables in that condition are

[9] A matrix representing the relative frequencies with which each of a number of stimuli is mistaken for each of the others by a person in a task requiring recognition or identification of stimuli.

seldom confused and this evidence clearly shows the complementary nature of bimodal speech (Figure 10).

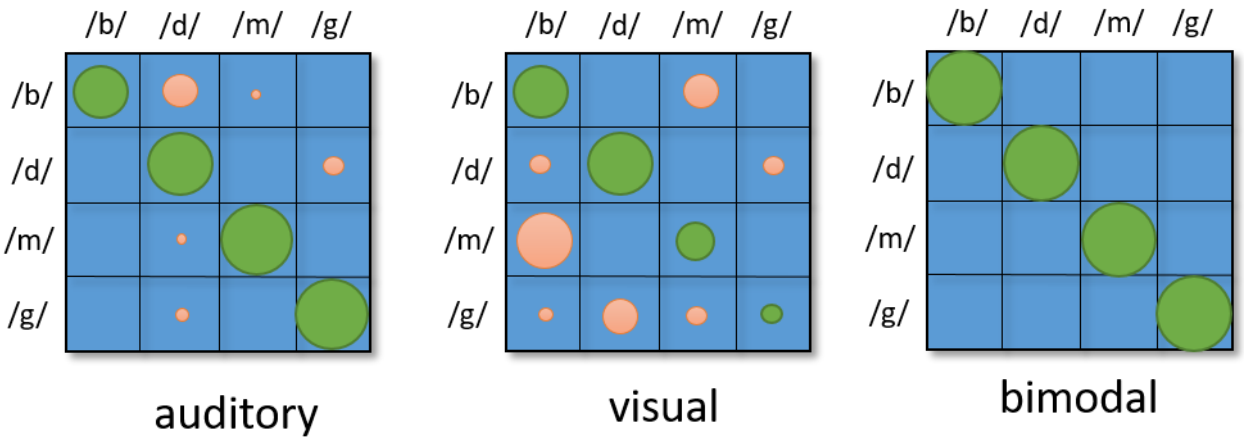


Figure 10 – Matrices of confusion show that confusability between some phonemes is more likely to occur than some others, in the visual and auditory modality (Massaro, 1998).

A Bayesian explanation for the McGurk effect

Massaro (1998) proposes that the underlying logic for this model is explained by the Bayesian statistical theorem that scientists often use to evaluate the predictive power of hypotheses. The mathematician Thomas Bayes (1702-1761) studied the problem of determining the probability of the causes through the observed effects. The theorem that bears his name refers to the probability of an event that is presented as the sum of several mutually exclusive events, therefore, when experimental subjects receive auditory and visual input, they must effective choose among several competing hypotheses, that is, they must posit an interpretation of their data. Each piece of information is evaluated with reference to a stored prototype to determine the degree to which the data support a given category. Auditory and visual probabilities are then integrated, and finally the relative goodness-of-match rule normalizes the results by comparing the support for each syllable to the combined scores for all.

Massaro (1998) applies the Bayesian theorem to speech data integration in the following way: “The probability that a perceived syllable falls into a speech category (c) given the acoustic evidence (A) is denoted P(c | A). We can state this probability in terms of the acoustic evidence given the category, the probability P(A | c), the probability of the category c, and the sum of the probabilities of observing all possible categories –In this case the total probability of finding the acoustic evidence A:

$$P(c \mid V) = \frac{P(A \mid c)P(c)}{sum_A}$$

The same logic holds for the probability of a category c given the visual evidence V:

$$P(c \mid V) = \frac{P(V \mid c)P(c)}{sum_V}$$

The desired probability given evidence from both modalities, P(c | A & V), also arises from Bayes's theorem. If A and V are conditionally independent –that is, if P(A & V | c) = P(A | c) P(V | c), Bayes's theorem can yield the optimal sensory-integration scheme” (p. 240):

$$\frac{P(c \mid A)P(c \mid V)P(c)}{sum_{AV}}$$

In the classic McGurk example, a subject is presented with incongruent information contrasting stimuli A (auditory /ba/) and V (visual /ga/). Each piece of information is evaluated with reference to a stored prototype to determine the degree to which the data support a given category (c) for instance, P(c I A), the probability the syllable fits category c given auditory information A. Auditory and visual probabilities are then integrated, and finally the relative

goodness-of-match rule normalizes the results by comparing the support for each syllable to the combined scores for all. This process shows how a subject might consider /da/ the best response in the classic example (Figure 11).

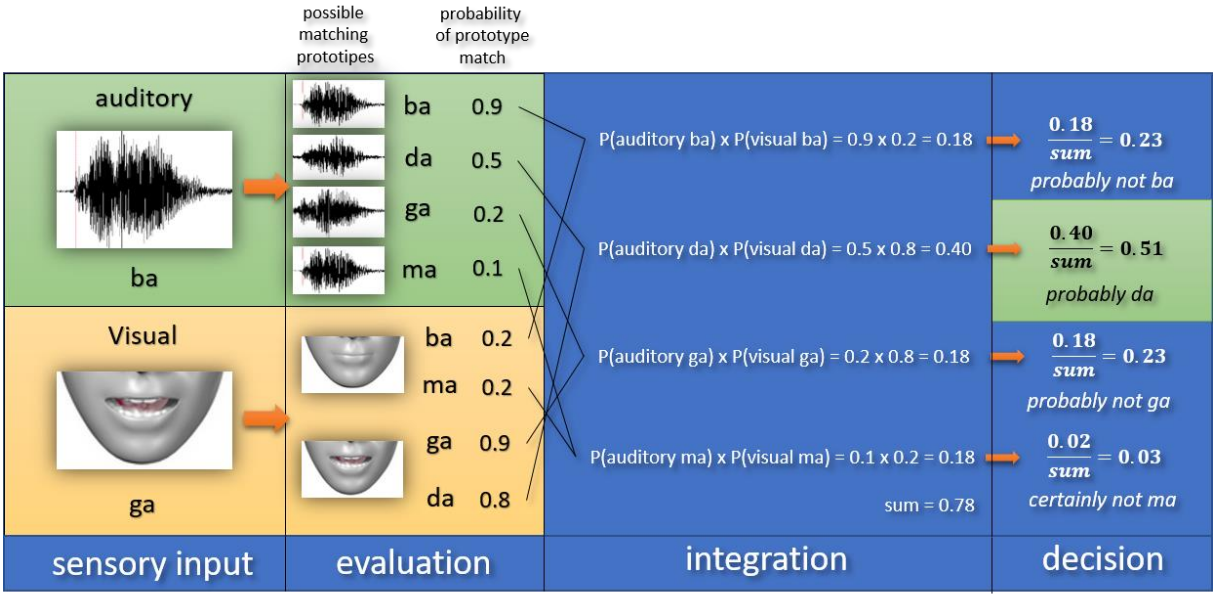


Figure 11 – The Bayesian reasoning for the McGurk effect (Massaro, 1998).

Weighted modalities for speech perception

Research has shown that not all incongruent audiovisual speech signals produce the same probability of a correct response, and that the brain appears to give greater weight to auditory information in some cases, and in others it relies heavily on the visual modality, and this weighting of the information seems to be based on its reliability, and later on integrated in a statistically optimal manner. Evidence supporting this claim comes from studies where it has been shown that when listeners are presented with degraded auditory input, this may contribute to an overreliance on the visual percept, causing visual information to dominate responses. Cienkowski and Carney (2002), for instance, reported that older adults with hearing loss tend

to respond with the visual portion of incongruent auditory–visual stimuli. Dodd et al. 2008 have also observed that children with hearing impairments appear to rely more on the visual signal compared to their typically developing peers and therefore report hearing the visually articulated consonants.

The upshot of this preceding discussion is that a listener's ability to weigh different sources of evidence might be shaped by sensory experience. As a direct consequence of this, an inquiry can be posed: Whether a listener's ability to weigh different modalities in audiovisual speech perception can be enhanced by early, extended and formal musical training, driving to a stronger weighting of acoustic cues present in the audiovisual speech stimuli.

Chapter 4 – Musical perceptual expertise

According to Vassena (2016), "Expert musicianship typically reflects skilled performance that constitutes one of the most complex human abilities, involving exposure to complex auditory and visual stimuli, memorization of elaborate sequences, and extensive motor rehearsal" (p. 6). In this chapter, we will discuss mainly three issues: (1) evidence for auditory enhancements elicited by musical training, (2) evidence supporting the claim that musical auditory training can have a positive effect in the encoding and recognition of acoustic cues relevant for speech perception and intelligibility; and (3) disputing evidence for the former that leads to the conclusion that speech and music are not directly linked and one processing system does not affect the other directly.

Auditory enhancements in musicians

Previous research has suggested that use-dependent functional reorganization extends across the sensory cortices to reflect the pattern of sensory input processed by a subject during development of musical skill (Pantev et al., 1998; 2001; Baumann et al., 2008). These results indicate that the effect of music expertise, which was traced by current density mapping to the auditory cortex, reflects an enlarged neuronal representation for specific sound features, such as *pitch* and *timbre*. Other recent work has shown that musical training sharpens the response of the brainstem to auditory stimuli and that this is more significant for children than adults (Tierney, Krizman, Skoe, Johnston, & Kraus, 2013).

Behavioral studies have shown that musicians perform better than non-musicians in both pitch discrimination (Spiegel & Watson, 1984; Tervaniemi, Just, Koelsch, Widmann, &

Schröger, 2005) and rhythmic performance (Drake, C., 1993; Habibi et al., 2014; Schaal et col., 2015; Klyn et col., 2016). This is consistent with the finding that musicians are more sensitive to some acoustic features critical for both speech and music processing (Spiegel and Watson 1984; Kishon-Rabin et al. 2001; Micheyl et al. 2006; Anderson and Kraus 2011). Structural imaging studies of the brain of musicians have reported increased size of corpus callosum (Schlaug et al., 1995), increased grey matter volume in motor cortex (Elbert et al., 1995), cerebellar regions (Hutchinson et al., 2003) and corticospinal tract (Imfeld et al., 2009). These areas are all considered to be directly involved in attaining musicals skills (Munte et al., 2002; Hannon and Trainor, 2007).

Early training

Skilled musicians often begin training early in life. Evidence for a sensitive period for musical training comes from studies showing that musicians who began lessons before age seven showed structural differences in the corpus callosum and sensorimotor cortex compared to those who began later (Amunts et al., 1997; Schlaug et al., 1995). More recent work has further addressed this question by controlling for any inherent differences in the length of training between musicians who begin training earlier than those who begin later. A series of behavioral and brain imaging studies compared early-trained (before age seven) and late-trained (after age seven) musicians who were matched for years of musical experience, years of formal training, and hours of current practice showed that early-trained musicians perform better on rhythm synchronization and melody discrimination tasks (Penhue et al., 2011; Vaquero et al., 2016; Steele et al., 2013, Krall, 2013) and have enhancements in gray- and white-matter structures in motor regions of the brain (Bailey et al., 2014; Steele et al., 2013). Gaser and

Schlaug (2003) showed evidence for structural brain changes after only 15 months of musical training in early childhood, which were correlated with improvements in musically relevant motor and auditory skills. Based on these findings, it can be argued that *early training* creates a behavioral and brain scaffold on which later practice can build (Penhune, 2011). In summary, literature discussed above has proposed strong evidence in favor of the claim that there is a correlation between musical training and some structural and functional changes in the brain.

Musical training as an enhancement for speech perception

In this section I will discuss some evidence supporting the claim *of a* possible *cross-domain auditory plasticity* –the possibility that training in one cognitive domain might affect the development of a different one, in this case, music training and speech perception.

Kraus and Chandrasekaran (2010) suggest that the neural encoding of speech can be enhanced by non-linguistic auditory training, and that musical training –learning to play a musical instrument or sing-, can have a positive effect in the encoding and recognition of acoustic cues relevant for speech perception and intelligibility. They point out that both music and speech use *pitch*, *timing*, and *timbre* to convey information, and suggest that years of processing these cues in a fine-grained way in music may enhance their processing in the context of speech. Along this line, Musacchia, Sams, Skoe, & Kraus (2007) showed that musicians had earlier and larger brainstem responses than non-musicians controls to both speech and music stimuli presented in auditory and audiovisual conditions, evident as early as 10 milliseconds after acoustic onset. Previous studies have also shown that musicians exhibit enhanced neural differentiation of stop consonants early in life and with as little as one year of training (Strait, O'Connell, Parbery-Clark, & Kraus, 2014). Musicians have also shown

perceptual improvements over non-musicians in both native and foreign linguistic domains (Magne et al., 2006; Marques et al., 2007; Moreno et al., 2009; Tzounopoulos and Kraus, 2009). Research has shown evidence that musicians perform better than non-musicians at segmenting speech from background noise (Parbery-Clark et al., 2013), pitch (Besson et al., 2007), and in prosodic tasks (Thompson et al., 2004). In summary, these studies suggest that music training improves the perception of minimal acoustic differences between auditory stimuli, and this might underlie language development.

The OPERA hypothesis

Patel has proposed the "OPERA" theory (2011; 2014) involving *cross-domain plasticity*, in which experience or training in one domain influences the neural processing in another domain. According to Patel, "musical training drives adaptive plasticity in speech processing networks when 5 conditions are met. These are: (1) Overlap: there is anatomical overlap in the brain networks that process an acoustic feature used in both music and speech (e.g., waveform periodicity, amplitude envelope); (2) Precision: music places higher demands on these shared networks than does speech, in terms of the precision of processing; (3) Emotion: the musical activities that engage this network elicit strong positive emotion; (4) Repetition: the musical activities that engage this network are frequently repeated, and (5) Attention: the musical activities that engage this network are associated with focused attention. According to the OPERA hypothesis, when these conditions are met neural plasticity drives the networks in question to function with higher precision than needed for ordinary speech communication." (p. 102). Nonetheless, Patel (2014) states that "conclusive proof that nonverbal instrumental musical training changes the neural processing of speech has yet to be obtained."

A "speech mode" for perception

The idea of a "speech mode" for perception proposes that speech is processed by special and unique cognitive and neural mechanisms and that these mechanisms are different than those used for processing other sounds. Evidence for this has been proposed by studies suggested the feasibility of cognitive and neural dissociations between musical and linguistic processing modules (Strait & Kraus, 2011; Besson, Chobert, & Marie, 2011). Evidence of patients with specific brain damage has shown that the ability to interpret speech can be profoundly impaired, yet the perception of musical sounds shows no impairment, or vice versa (Peretz, 1993; Peretz et al., 1994; Ayotte et col., 2000; Ayotte et al., 2002). Evidence for the separability of speech and music is demonstrated by the condition known as *Auditory Verbal Agnosia* or *Pure Word Deafness* (Pure Word Deafness | ScienceDirect Topics., n.d.), an exceptionally rare and specific type of auditory agnosia, whose primary symptom is the inability to comprehend spoken words, while the abilities to speak, read, and write remain intact. Crucially, the ability to process non-speech auditory information, including music, also remains relatively unaltered (Poeppel, 2001). The existence of this clinical condition demonstrates a dissociation between speech perception, non-speech auditory processing, and central language processing. Poeppel (2001) argues that "speech perception and non-speech recognition may function in parallel. (…) The systems rely on shared mechanisms early in the processing stream but make use of specialized architecture in the construction of the relevant percepts. Moreover, they can be damaged selectively" (p. 680).

Previous related research using the McGurk Effect

Proverbio, Massetti, Rizzi, and Zani (2016) carried out a study in which musicians (N = 40) and non-musicians (N = 40) were tested in a McGurk task. Their results across groups and conditions (auditory vs incongruent audiovisual condition, collapsed across phonemes), gave rise to the significance of group factor ($F_{2,14} = 6.3$, $p = 0.03$), with musicians showing a better performance than controls. No significant decrease in performance between the congruent and the incongruent McGurk conditions was found in musicians.

Their paper does not mention the native language of participants, although it mentions that one of their aims was to investigate in greater depth the existence of the McGurk effect in Italian. Evidence for cross-language difference in McGurk effect affectation has already been discussed in this thesis. In order to tackle this issue, in my experiment, all the participants (musicians and non-musicians) and also the recorded voices are native English speakers and non-bilinguals. I have no knowledge of another similar study carried out in English.

Another important aspect is the age of starting musical training. Whereas Proverbio et al. used a group of experts that show a wide range of starting age (range: 4 – 18, M = 8.125), we decided to use only participants with early training, started between the age of 3 and 9, based on the previous literature review arguing that musicians who began training around the age of 6 show better task performance and greater changes in auditory and motor regions of the brain.

Chapter 5 – Current experiment

Hypothesis

I hypothesized that early, extended and formal musical training can have a significant effect in weighting acoustic cues present in the audiovisual speech signal, therefore, music experts should show higher accuracy over controls identifying the mismatching auditory component in a McGurk paradigm.

Participants

A total of 81 participants took part in this study, 41 music students (age $\bar{x}$ = 21.8) at undergraduate and graduate levels were recruited from the School of Music at the University of Victoria (BC, Canada); 13 pianists, 4 woodwinds, 7 brasses, 3 strings, 10 singers and 4 guitarists. They all started training between the age of 3 and 9 (age started $\bar{x}$ = 6.57), and have spent more than 10 years formally practicing music (years of practice $\bar{x}$ = 15.8). All reported having no auditory impairments. As control group, 40 individuals (age $\bar{x}$ = 20.24) were recruited using the SONA system of the Department of Psychology of the University of Victoria (BC, Canada). They reported not having any auditory impairment, and not having any type of musical instrumental or vocal training apart from regular exposure to music class at public school (without musical profile/orientation). All reported to be native English speakers and non-bilinguals. All the participants signed a consent form to take part of this study.

Stimuli

Stimuli consisted of audiovisual recordings of two female and two male native English speakers (one pair used for practice trials and one pair used for critical trials) pronouncing 4 syllables: /ba/, /ga/, /pa/ and /ka/. For the recordings, a Canon Rebel X3 camera (MPEG4 24-bits, 1680 x 1085-pixel resolution) was used together with a ZOOM H4n hand recorded (PCM 24 bits, 48000 Khz), in an isolated room, with artificial light. The software used for editing was Adobe Premiere Pro CC 2017. Video was cropped using a grid centered on the position of the lips, leaving just the lower part of the face, with enough space to allow vocal movement without exceding the crop borders. For the visual presentation, the original size of the videos was maintained (1680 x 1085-pixel). The grid was removed from the final presentation of stimulus videos.

I used a CV (consonant-vowel) structure in which the vowel is always /a/, and the consonants were four *plosives*: Two *voiced* ([b] and [g]) and two *unvoiced* ([p] and [k]). (See *appendix 1* for the complete set of spectrograms of the auditory stimuli used). The criteria for aligning the timing in view of the different VOT (voice onset time) and speech lip movements in the incongruent syllables, onset of the sonority point of the utterance was detected using Praat 6.0.42 software, as a mark point after VOT in the auditory track. When replacing the congruent sound with incongruent, these markers were aligned. The red dashed lines in the spectrograms (the full set of spectrograms of the syllables used for this study are presented in appendix 3) indicate the moment of the start of the sonority. The time between vertical lines corresponds to the duration of the VOT. To measure the VOT, the usual procedure is to consider the explosion as a point of reference and depending on it measure the temporal distance from this point to the

beginning of the sonority. Vocalization started at 600 ms after the video onset and continued for approximately 500 ms. The total duration of each video is 1500 milliseconds. The audio gain was normalized to -10 decibels relative to full scale at the voiced peak level for all stimuli (usually 0 dBFS is the loudest level allowed in a digital system), meaning that all stimuli would be perceived as having the same intensity. For this, the intensity peak that occurs when the vowel is articulated was taken as reference for normalization.

The set of stimuli used -32 videos in total- was arranged in 4 conditions: ***Congruent*** (matching audio and visual), ***Incongruent*** (mismatching audio and visual) ***Auditory only*** and ***Video only***. In order to create the incongruent stimuli, the original audio from a syllable was replaced by the audio of another syllable, from the same set of videos. For the ***visual only*** condition, the audio was removed, leaving only the visual, and for the ***auditory only*** condition, the video was removed, leaving only the audio with a blank screen.

Condition	**Stimuli (V: visual – A: audio)**
Congruent	[V:/ba/ - A:/ba/] [V:/ga/ - A:/ga/] [V:/pa/ - A:/pa/] [V:/ka/ - A:/ka/]
Incongruent	[V:/ba/ - A:/ga/] [V:/pa/ - A:/ka/] [V:/ga/ - A:/ba/] [V:/ka/ - A:/pa/]
Auditory only	[V:/X/ - A:/ga/] [V:/X/ - A:/ba/] [V:/X/ - A:/ka/] [V:/X/ - A:/ka/]
Visual only	[V:/ba/ - A:/X/] [V:/ga/ - A:/X/] [V:/pa/ - A:/X/][V:/pa/ - A:/X/]

Table 2 – Set of stimuli used for this study.

Procedure

The experimental procedure was designed using the software E-Prime 2.0. Participants were required to watch the videos on a computer screen (Dell Optiflex 740) using headphones (Sennheiser HD210 connected to a Roland RUBIX 24 digital audio interface) and respond to what they hear, using a computer keyboard with 8 labeled keys. A 2-second fixation cross is set at the center of the lips, followed by the video. A cue for entering the response is presented after each stimulus. Participants completed a total of 96 trials (32 practice trials and 64 critical trials, divided in sets of 16 separated by a pause break), with no feedback for correct or incorrect responses. For the practice trials, different faces were used and the possibility of adjusting the volume of the headphones was offered. Participants have a time window of 5 seconds after the cue for response input. After that time, the response is considered 'unknown – no response'. The *visual only* condition served as a way of preventing participants from looking away from the computer monitor, although they were also being supervised by the researcher while they carried out the experimental task, in order to prevent any disruption of the procedure. As in this study I am interested in auditory responses, the results for the visual only condition are not included in the statistical analysis.

Results

The mean of accuracy was computed for each condition with the two groups and those means are shown in Figure 12. Air bars show confidence intervals. The effect of condition and musical training were assessed in a repeated measures two-factor analysis of variance

(ANOVA) with type I error rate set at .05. The ANOVA revealed a main effect for ***condition***: $F(2, 160) = 235.9$, $p < 0.001$. There was no significant ***group*** effect: $F(1, 80) = 0.39$, $p = 0.52$.

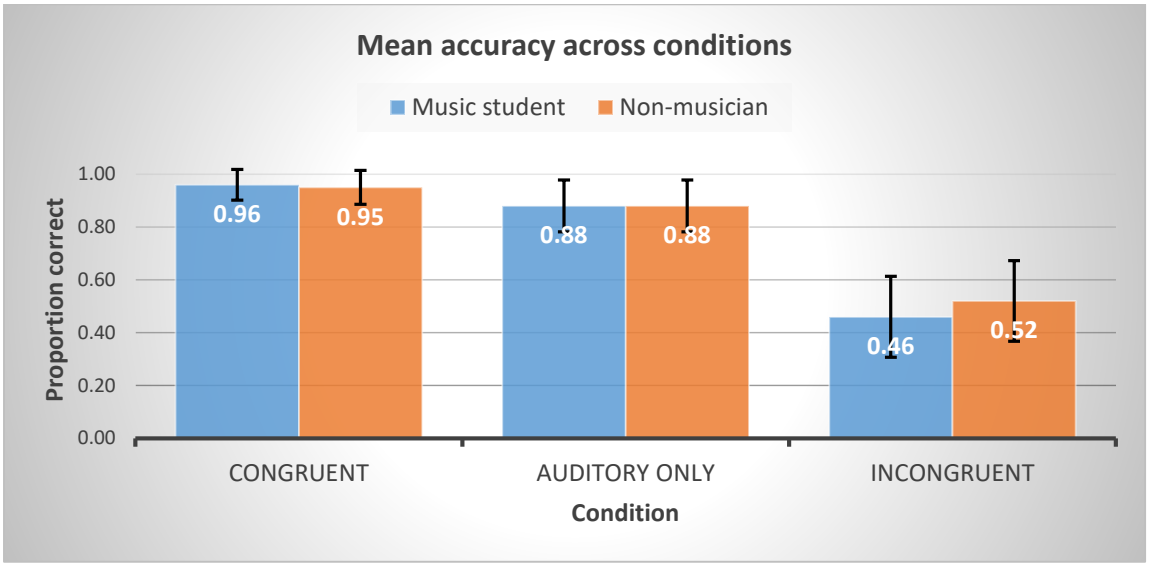


Figure 12 – Accuracy mean for congruent, neutral and incongruent conditions for both groups.

The compared accuracy between conditions collapsing the groups (Fig. 13) showed an interaction with ***sounds***, $F(6, 480) = 31.21$, $p < 0.001$, especially noticeable for /pa/.

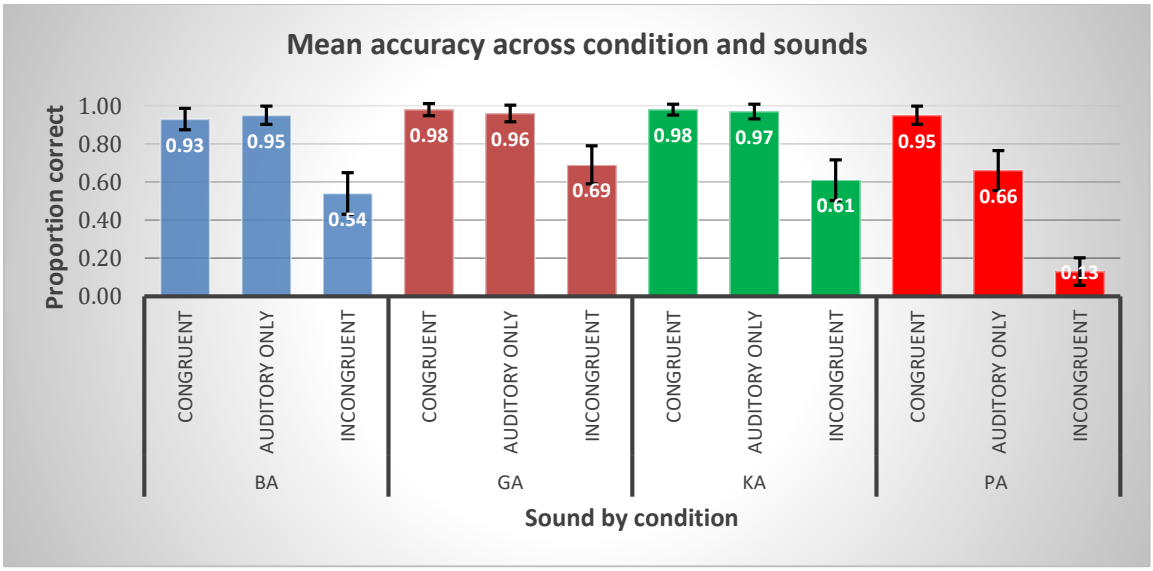


Figure 13 – Proportion correct for sounds across conditions with collapsed groups.

Significant accuracy differences were observed for each of the used ***speakers*** (male and female recorded voices) $F(1,80) = 50.55$, $p < 0.001$.. An interaction between ***speaker*** and ***condition*** was observed, showing that the accuracy mean for each condition was affected by the perceived *speaker*. $F(2, 160) = 30.81$, $p < 0.001$. (Fig. 14)

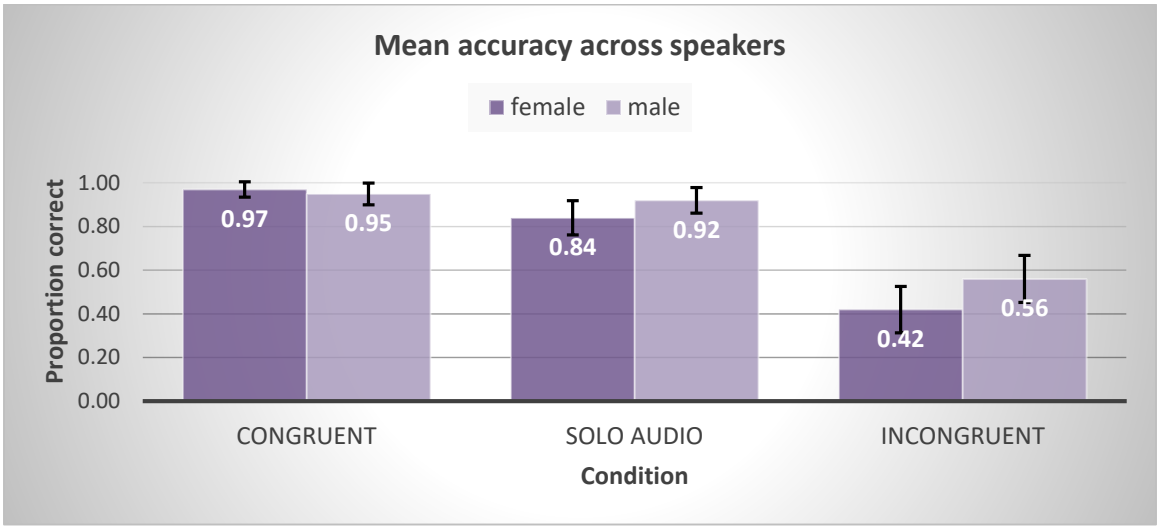


Figure 14 – Accuracy mean for congruent, neutral and incongruent conditions for each speaker voice.

Discussion

Our data show a large amount of inter-subject variability in the extent to which individual participants are susceptible to the McGurk illusion, present in most of the literature, with individuals in both groups reporting incorrect responses ranging from 0 to 100%. The compared accuracy across conditions shows clear evidence that there is a significant perceptual enhancement produced by the sight of the congruent lip movements of the speaker comparing congruent vs. solo audio, and a significant worse performance when the visual information is incongruent with the auditory. This is more evident for some sounds than others: The sound /pa/ shows an astonishing decreasing performance of .82 (mean of proportion correct) from the

congruent (.95) to the incongruent (.13). These results corroborate previous findings (Sumby & Pollack, 1954; Sommers et. al., 2016) supporting the idea that visual information heavily influences speech perception.

In this study, the mean of accuracy was significantly higher for the male voice than for the female and also an interaction between voice and conditions was observed. I did not find any previous study accounting for voice gender differences in McGurk perception. Other studies have shown a reduction of the effect by presenting incongruent voice gender for the sound and the visual (Chan, Notbohm, Naumer, Bosch, & Kaiser, 2013), but I did not find a large-scale study accounting for inter-gender differences. As I only used one sample voice for each gender, it would not be valid to attribute this difference to a gender effect, and it might be driven by inter-stimulus differences. However, the research in this line of inquiry is scarce, and an extended study looking to account for voice gender differences in McGurk perception might show interesting results.

Previous research has proposed that an individual's susceptibility to the McGurk effect might be correlated to their ability to extract place of articulation information from the visual signal (i.e., a more fine-grained analysis of lip reading ability), but not to their scores on tasks measuring attentional control, processing speed, working memory capacity, or auditory perceptual gradiency (Brown et al., 2018; Witten and Knudsen, 2005). In my study I found a trend of lower accuracy in the incongruent condition –although not significant- for *singers* within the group of musicians. A post-hoc analysis of statistical power for a sample size of 10 participants for this subgroup gave a result of 24%, I hypothesize that singers might show a worse performance in the incongruent condition than other instrumentalists, due to a recurring

exposure to facial gestures and expressions, as part of their specific vocal training, that might derive in a better ability to extract the place of articulation compared to other musicians, positively correlated with McGurk sensitivity.

I believe that future research aiming towards finding group differences *within* musicians might show interesting results, regarding to auditory enhancements provoked by a specific type of auditory training that can vary between different types or families of instruments and singers. For this, I consider that a more refined and sensitive version of the McGurk paradigm is needed.

Conclusion

The first and foremost finding for this study is the absence of significant group differences between native English speakers with early, extended and formal musical training versus individuals without it. These results suggest that music training does not prevent experiencing the McGurk illusion to a significant extent, and that the weighting of auditory cues for speech perception might not be affected by musical training. These findings propose no evidence to support the claim that music training favours the encoding and processing of speech, and also failed to replicate Proverbio et al. (2016) findings. This data is consistent with the claim that language and music are separable cognitive domains, where the perception of speech might be unaffected by extensive musical training.

The causes for the large extent of inaccurate identification of the auditory component and the wide inter-subject variability in McGurk susceptibility are still not well understood, along with many other areas of inquiry, such as the nature and structure of auditory memory. In the future, I am interested in conducting research that aims to further understand how musical ability relates to other cognitive abilities such as speech and general intelligence, how the brain structures and cognitive systems involved in music processing and performance interact with one another, and specially, how these interactions change over time due to musical training, with the principal goal of better understanding the link between brain and behavior in musical perceptual expertise.

Bibliography

Altieri, N. & Yang, C. (2016) Parallel linear dynamic models can mimic the McGurk effect in clinical populations. *Journal of Computational Neuroscience*. 41(2), 143-155. doi: 10.1007/s10827-016-0610-z

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125*(2), 238–251. https://doi.org/10.1093/brain/awf028

Ayotte, J., Peretz, I., Rousseau, I., Bard, C., & Bojanowski, M. (2000). Patterns of music agnosia associated with middle cerebral artery infarcts. *Brain*, *123*(9), 1926–1938. https://doi.org/10.1093/brain/123.9.1926

Baumann, S., Meyer, M., & Jancke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience,* 20, 2238–2249. doi: 10.1162/jocn.2008.20157.

Bear, H., Harvey, R. (2017) Phoneme-to-viseme mappings: the good, the bad, and the ugly. *Speech Communication*. 95, 40-67. doi: 10.1016/j.specom.2017.07.001.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention and memory. *Frontiers in Psychology*, *2*. Retrieved from https://doaj.org

Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: experience-dependent effects in the double-flash illusion. *Experimental brain research*, *234*(10), 3037-3047. doi: 10.1007/s00221-016-4705-6

Brown, R. M., Zatorre, R. J., & Penhune, V. B. (2015). Chapter 4 - Expert music performance: Cognitive, neural, and developmental bases. In E. Altenmüller, S. Finger, & F. Boller (Eds.), *Progress in Brain Research* (pp. 57–86). https://doi.org/10.1016/bs.pbr.2014.11.021

Chan, J., Notbohm, A., Naumer, M., van den Bosch, J., & Kaiser, J. (2013). The influence of gender incongruence on the McGurk-percept: A combined behavioural and fMRI study. *Multisensory Research*, *26*(1–2), 184–185. https://doi.org/10.1163/22134808-000S0138

Colman, A. (2015). Speech spectrogram. In A Dictionary of Psychology. : Oxford University Press,. Retrieved 24 Jun. 2019, from https://www-oxfordreference-com.ezproxy.library.uvic.ca/view/10.1093/acref/9780199657681.001.0001/acref-9780199657681-e-7871.

Colman, A. M. (2015). Formant. In *A Dictionary of Psychology*. Retrieved from http://www.oxfordreference.com/view/10.1093/acref/9780199657681.001.0001/acref-9780199657681-e-3256

Colman, A. (2015). Confusion matrix. In A Dictionary of Psychology. : Oxford University Press,. Retrieved 26 Jun. 2019, from https://www-oxfordreference-com.ezproxy.library.uvic.ca/view/10.1093/acref/9780199657681.001.0001/acref-9780199657681-e-1778.

Davies, S. (2014). Music, Definitions of. In W. Thompson, *Music in the Social and Behavioral Sciences: An Encyclopedia*. https://doi.org/10.4135/9781452283012.n245

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics*, 53(1), 25–33. doi: 10.3758/BF03211712.

Eidels, A., Houpt, J., Altieri, N., Pei, L., & Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: a theory of interactive parallel processing. *Journal of Mathematical Psychology*, 55(2), 176–190. doi: 10.1016/j.jmp.2010.11.003.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B. & Taub, E. (1995) Increased cortical representation of the fingers of the left hand in string players. *Science*, 270, 305–307. doi: 10.1126/science.270.5234.305

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Fisher, C. (1968). Confusions among visually perceived consonants. *Journal of Speech, Language, and Hearing Research*, *11*(4), 796-804. doi: 10.1044/jshr.1104.796

Fowler, C., & Saltzman, E. (1993). Coordination and Coarticulation in Speech Production. *Language and Speech*, 36(2–3), 171–195. doi: 10.1177/002383099303600304

Fromkin, V. A. (2014). *Tone: A Linguistic Survey*. Academic Press.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., Trainor, L.J., (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, 129, 2593–2608. doi: 10.1093/brain/awl247.

Ganesh, A., Berthommier, F. & Schwartz, J. (2018). Audiovisual Binding for Speech Perception in Noise and in Aging. *Language Learning*, 68, 193-220. doi:10.1111/lang.12271

Gaser, C. & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. The *Journal of Neuroscience*, 23(27), 9240-9245 doi: 10.1523/jneurosci.23-27-09240.2003.

Green, K., Gerdeman, A. (1995). Cross-modal discrepancies in coarticulation and the integration of speech information: The McGurk effect with mismatched vowels. Journal of Experimental Psychology: *Human Perception and Performance*, 21(6): 1409–1426. doi: 10.1037/0096-1523.21.6.1409

Greenberg, S. (2006). A multi-tier framework for understanding spoken language. In Greenberg & W. A. Ainsworth, Eds., Listening to speech: An auditory perspective (p. 411). Mahwah, NJ. Lawrence Erlbaum Associates. ISBN: 0805845399.

Gussenhoven, C., & Jacobs, H. (2005). *Understanding phonology* (2nd ed). London: Hodder Arnold.

Habibi, A., Wirantana, V., & Starr, A. (2014). Cortical activity during perception of musical rhythm: Comparing musicians and nonmusicians. *Psychomusicology*, 24(2), 125-135. URL:http://search.proquest.com.ezproxy.library.uvic.ca/docview/1549612499?accountid=14846

Hannon, E., & Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, 11(11), 466-472. doi: 10.1016/j.tics.2007.08.008.

Hellbernd, N. & Sammler, D. (2016) Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language.* 88, 70-86, ISSN 0749-596X, https://doi.org/10.1016/j.jml.2016.01.001.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, *97*(5), 3099-3111. doi: 10.1121/1.411872

Hisanaga, S., Sekiyama, K., Igasaki, T. & Murayama, N. (2016) Language/Culture Modulates Brain and Gaze Processes in Audiovisual Speech Perception. *Scientific Reports*, 6, 35265. doi: 10.1038/srep35265.

Hutchinson, S., Hui-Lin Lee, L., Gaab, N. & Gottfried Schlaug (2003) Cerebellar Volume of Musicians. *Cerebral Cortex*, 13(9) 943–949. doi: 10.1093/cercor/13.9.943

Imfeld, A., Oechslin, M. S., Meyer, M., Loenneker, T., & Jancke, L. (2009). White matter plasticity in the corticospinal tract of musicians: A diffusion tensor imaging study. *Neuroimage,* 46, 600–607. doi: 10.1016/j.neuroimage.2009.02.025.

International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press. ISBN 0-521-65236-7.

Jaekel, B. N., Newman, R. S., & Goupell, M. J. (2017). Speech Rate Normalization and Phonemic Boundary Perception in Cochlear-Implant Users. *Journal of Speech, Language, and Hearing Research : JSLHR*, *60*(5), 1398–1416. https://doi.org/10.1044/2016_JSLHR-H-15-0427

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology*, *29*(1), 1–23. https://doi.org/10.1006/cogp.1995.1010

Klatt, D. H., Klatt, L. C., (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustic Society of America.* 87(2). 820-857. doi: 10.1121/1.398894.

Kouroupetroglou, G., & Chrysochoidis, G. (2014, June). Formant tuning in Byzantine chanting. In 1st International Interdisciplinary Conference "The Psaltic Art as an Autonomous Science", Volos, Greece.

Krall, A. (2013). Auditory critical periods: A review from system's perspective. *Neuroscience*, 247, 117-133. doi: 10.1016/j.neuroscience.2013.05.021.

Krampe, R. Th., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General*, *125*(4), 331–359. https://doi.org/10.1037/0096-3445.125.4.331

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Ladefogued, P. (1996) Elements of acoustic phonetics. University of Chicago Press. Chicago, IL. ISBN: 02264676350226467635.

Lakatos, P. et al. (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, 53, 279–292.

Law, J. L., & Rennie, R. R. (2015). Pitch. In J. Law & R. Rennie (Eds.), *A Dictionary of Physics*. Retrieved from http://www.oxfordreference.com/view/10.1093/acref/9780198714743.001.0001/acref-9780198714743-e-2339

Lee, H. L., Noppeney, U. (2014). Music expertise shapes audio visual temporal integration windows for speech, sine wave speech and music. Frontiers of Psychology, 5, 868. doi: 10.3389/fpsyg.2014.00868.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Lieberman, P. (1991). Uniquely human: The evolution of speech, thought, and selfless behavior. Cambridge, Mass: Harvard University Press. ISBN: 0674921828.

Magne, C., Schon, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than non-musician children: behavioral and electrophysiological approaches. *J. Cogn. Neurosci*., 18, 199–211. doi: 10.1162/089892906775783660

Magnotti, J. F., Basu Mallick, D., Feng, G., Zhou, B., Zhou, W., & Beauchamp, M. S. (2015). Similar frequency of the McGurk effect in large samples of native Mandarin Chinese and American English speakers. *Experimental Brain Research, 233*(9), 2581–2586. https://doi.org/10.1007/s00221-015-4324-7

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci*., 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453.

Massaro D., Cohen M., Tabain M., & Beskow J. (2012). Animated speech: research progress and applications. Audiovisual Speech Processing. Cambridge, MA. pp. 246–272. doi: 10.1017/cbo9780511843891.014.

Massaro, D. W. & Stork, D. G. (1998). Speech recognition and sensory integration: a 240-year-old theorem helps explain how people and machines can integrate auditory and visual information to understand speech. *American Scientist, 86*(3), 236-244. Retrieved from http://www.jstor.org/stable/27857023

Massaro, D. W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In *Categorical perception: The groundwork of cognition* (pp. 254–283). New York, NY, US: Cambridge University Press.

Massaro, D.W. (2001) Speech Perception. International Encyclopedia of the Social & Behavioral Sciences, (Pergamon, Italy) 14870-14875. ISBN 9780080430768 doi: 10.1016/B0-08-043076-7/01465-0.
http://www.sciencedirect.com/science/article/pii/B0080430767014650

Matthews, P. (2014). Voice onset time. In The Concise Oxford Dictionary of Linguistics. : Oxford University Press,. Retrieved 24 Jun. 2019, from https://www-oxfordreference-com.ezproxy.library.uvic.ca/view/10.1093/acref/9780199675128.001.0001/acref-9780199675128-e-3623.

Matthews, P. (2014). Co-articulation. In The Concise Oxford Dictionary of Linguistics. : Oxford University Press,. Retrieved 12 Dec. 2018, from http://www.oxfordreference.com.ezproxy.library.uvic.ca/view/10.1093/acref/9780199675128.001.0001/acref-9780199675128-e-549.

Maurer, D. (2016) Acoustics of the Vowel – Preliminaries. Peter Lang International Academic Publishers. ISBN 978-3-0343-2031-3. Retrieved from http://www.oapen.org/record/588650

McArthur, T., Lam-McArthur, J., & Fontaine, L. (2018). Vowel. In (Ed.), The Oxford Companion to the English Language. : Oxford University Press,. Retrieved 10 Nov. 2018, from http://www.oxfordreference.com.ezproxy.library.uvic.ca/view/10.1093/acref/9780199661282.001.0001/acref-9780199661282-e-1287.

McArthur, T., Lam-McArthur, J., & Fontaine, L. (2018). Consonant. In (Ed.), The Oxford Companion to the English Language. : Oxford University Press,. Retrieved 12 Dec. 2018, from http://www.oxfordreference.com.ezproxy.library.uvic.ca/view/10.1093/acref/9780199661282.001.0001/acref-9780199661282-e-306.

McArthur, T., Lam-McArthur, J., & Fontaine, L. (2018). Language. In The Oxford Companion to the English Language. : Oxford University Press,. Retrieved 9 Feb. 2019 from http://www.oxfordreference.com.ezproxy.library.uvic.ca/view/10.1093/acref/9780199661282.001.0001/acref-9780199661282-e-682.

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746. https://doi.org/10.1038/264746a0

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex.*, 19, 712–723. doi: 10.1093/cercor/bhn120.

Morís Fernández, L., Macaluso, E., & Soto-Faraco, S. (2017). Audiovisual integration as conflict resolution: The conflict of the McGurk illusion. *Hum. Brain Mapp.*, 38, 5691-5705. doi: 10.1002/hbm.23758.

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice Does Not Make Perfect: No Causal Effect of Music Practice on Music Ability. *Psychological Science*, *25*(9), 1795–1803. https://doi.org/10.1177/0956797614541990

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, 3(6): 473–478. Retrieved from http://ezproxy.library.uvic.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mnh&AN=12042882&site=ehost-live&scope=site

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. https://doi.org/10.1073/pnas.0701498104

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research, 241*(1), 34–42. https://doi.org/10.1016/j.heares.2008.04.013

Niall A.M. Klyn, Udo Will, Yong-Jeon Cheong, Erin T. Allen. (2016) Differential short-term memorisation for vocal and instrumental rhythms. *Memory,* 24(6), 766-791. doi: 10.1080/09298215.2014.937724.

Nygaard, L. C., & Pisoni, D. B. (1995). Chapter 3 - Speech Perception: New Directions in Research and Theory. In J. L. Miller & P. D. Eimas (Eds.), *Speech, Language, and Communication (Second Edition)* (pp. 63–96). https://doi.org/10.1016/B978-012497770-9.50005-4

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*(3), 172–191. https://doi.org/10.1037/0033-295X.85.3.172

Omata, K., Mogi, K. (2007). Fusion and combination in audiovisual integration. *Proc. R. Soc. A.* 464, 319-340. https://doi.org/10.1098/rspa.2007.1910.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*(6678), 811. doi: https://doi.org/10.1038/33918.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A. & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport, 12 (1), 169-174.

Patel, A. (2008) Music, language and the brain. Oxford University Press, New York, NY. ISBN 978-0-19-975530-1.

Patel, A. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology,* 2(142). https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research, 308,* 98–108. https://doi.org/10.1016/j.heares.2013.08.011

Penhune, V. (2011). Sensitive periods in human development: Evidence from musical training. *Cortex,* 47(9) 1126-1137 https://doi.org/10.1016/j.cortex.2011.05.010.

Peretz, I. (1993). Auditory atonalia for melodies. *Cognitive Neuropsychology*, *10*(1), 21–56. https://doi.org/10.1080/02643299308253455

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691. (12830160).

Peretz, I., Belleville, S., & Fontaine, S. (1997). Dissociations entre musique et langage après atteinte cérébrale: Un nouveau cas d'amusie sans aphasie. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *51*(4), 354.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994) Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117*(6), 1283-1301. https://doi.org/10.1093/brain/117.6.1283.

Peretz, I., Vuvan, D., Lagrois, M.-É., and Armony, J. L. (2015). Neural overlap in processing music and speech. Philol. Trans. R. Soc. Lond. B Biol. Sci. 370:20140090. https://doi.org/10.1098/rstb.2014.0090.

Pierrehumbert, J. B. (1987). *The phonology and phonetics of English intonation*. Indiana University Linguistics Club, Bloomington.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science, 25*(5), 679–693. https://doi.org/10.1207/s15516709cog2505_3

Proverbio, A. M., Massetti, G., Rizzi, E., & Zani, A. (2016). Skilled musicians are not subject to the McGurk effect. *Scientific Reports*, *6*(1), 30423. https://doi.org/10.1038/srep30423

Pure Word Deafness - an overview | ScienceDirect Topics. (n.d.). Retrieved June 25, 2019, from https://www.sciencedirect.com/topics/psychology/pure-word-deafness

Redford, M. A. (2015). *The Handbook of Speech Production*. John Wiley & Sons.

Schaal, N., Banissy, M., & Lange, K. (2015). The Rhythm Span Task: Comparing Memory Capacity for Musical Rhythms in Musicians and Non-Musicians. *Journal of New Music Research*, *44*(1), 3–10. https://doi.org/10.1080/09298215.2014.937724

Schlaug, G., Lutz Jäncke, L., Huang, Y., Staiger, J. & Steinmetz, H. (1995) Increased corpus callosum size in musicians. *Neuropsychologia*, 33(8): 1047-1055. https://doi.org/10.1016/0028-3932(95)00045-5.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., ... & Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature neuroscience, 8*(9), 1241. https://doi.org/10.1038/nn1530

Schnupp, J., Israel Nelken, I. & King, A. (2001). Auditory neuroscience: Making sense of sound. The MIT Press. Cambridge, MA. ISBN 978-0-262-11318-2

Schroeder, C., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008) Neuronal oscillations and visual amplification of speech. Trends in Cognitive Sciences, 12(3), 106-113. https://doi.org/10.1016/j.tics.2008.01.002.

Sekiyama, K. & Tohkura, Y. (1991) McGurk effect in non-English listeners: few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America*, 90(4), 1797-1805. https://doi.org/10.1121/1.401660.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. Percept. *Psychophys*, 59, 73–80. https://doi.org/10.3758/BF03206849.

Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. G. (1996). The role of practice in the development of performing musicians. *British Journal of Psychology, 87*(2), 287–309. https://doi.org/10.1111/j.2044-8295.1996.tb02591.x

Sommers, M. S., & Phelps, D. (2016). Listening effort in younger and older adults: A comparison of auditory-only and auditory-visual presentations. *Ear and Hearing*, *37*(Suppl 1), 62S-68S. https://doi.org/10.1097/AUD.0000000000000322

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *76*(6), 1690–1695. https://doi.org/10.1121/1.391605

Steele, C., Bailey, J., Zatorre, R. & Penhune, V. (2013). Early Musical Training and White-Matter Plasticity in the Corpus Callosum: Evidence for a Sensitive Period. *Journal of Neuroscience,* 33(3), 1282-1290; https://doi.org/10.1523/jneurosci.3578-12.2013

Stein, B.E. & Meredith, M.A. (1993) The merging of the senses. MIT Press.

Stevens, K. (1998) Acoustic phonetics. Cambridge, MA. MIT Press.

Stevens, S. S., & Volkmann, J. (1940). The Relation of Pitch to Frequency: A Revised Scale. *The American Journal of Psychology*, *53*(3), 329–353. https://doi.org/10.2307/1417526

Strait, D., & Kraus, N. (2011). Playing Music for a Smarter Ear: Cognitive, Perceptual and Neurobiological Evidence. *Music Perception: An Interdisciplinary Journal*, *29*(2), 133–146. https://doi.org/10.1525/mp.2011.29.2.133

Strait, D., O'Connell, S., Parbery-Clark, A., & Nina Kraus. (2014) Musicians' Enhanced Neural Differentiation of Speech Sounds Arises Early in Life: Developmental Evidence from Ages 3 to 30, *Cerebral Cortex,* 24(9), 2512–2521. https://doi.org/10.1093/cercor/bht103

Sumbey, W., & Pollack, I. (1954) Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.,* 26, 212-215 https://doi.org/10.1121/1.1907309.

Syrdal, A., & Gopal, S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America*, *79*(4), 1086–1100. https://doi.org/10.1121/1.393381

Tatham, M., & Morton, K. (2006). Speech production and perception. New York; Houndmills, Basingstoke, Hampshire; Palgrave Macmillan. https://doi.org/10.1057/9780230513969

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, *161*(1), 1–10. (15551089).

Tierney, A., Krizman, J., Skoe, E., Johnston, K., & Kraus, N. (2013). High school music classes enhance the neural processing of speech. *Frontiers in Psychology*, *4*, 855. https://doi.org/10.3389/fpsyg.2013.00855

Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00725

Titze, I. (2008). The Human Instrument. *Scientific American,298*(1), 94-101. Retrieved from http://www.jstor.org.ezproxy.library.uvic.ca/stable/26000381

Vaquero, L., Hartmann, K., Ripollés, P., Rojo, N., Sierpowska, J., François, C., ... & Münte, T. F. (2016). Structural neuroplasticity in expert pianists depends on the age of musical training onset. *Neuroimage*, *126*, 106-119. https://doi.org/10.1016/j.neuroimage.2015.11.008

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*(2), 192–212. https://doi.org/10.1037/0033-2909.101.2.192

Wallmark, Z. (2014). Timbre. In W. Thompson, *Music in the Social and Behavioral Sciences: An Encyclopedia*. https://doi.org/10.4135/9781452283012.n392

Weeks, K. (2019). *The SAGE Encyclopedia of Lifespan Human Development* (By pages 2126-2127). https://doi.org/10.4135/9781506307633

Williams, J. J., Rutledge, J. C., Garstecki, D. C., & Katsaggelos, A. K. (1997). Frame rate and viseme analysis for multimedia applications. *Proceedings of First Signal Processing Society Workshop on Multimedia Signal Processing*, 13–18. https://doi.org/10.1109/MMSP.1997.602606

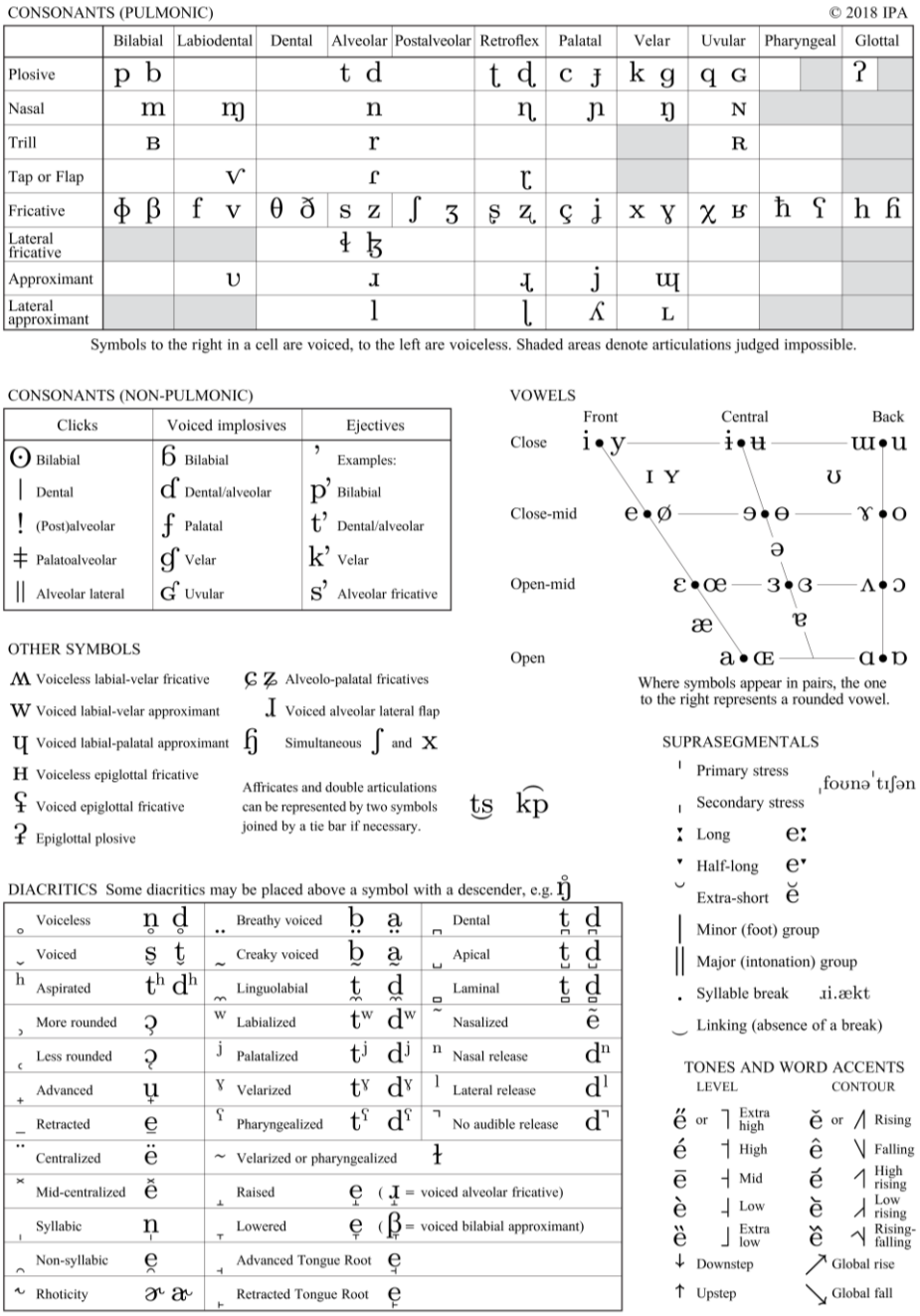
Wingfield, A. & Lash, A. (2016) Audition and Language Comprehension in Adult Aging: Stability in the Face of Change - Chapter 9 - *Handbook of the Psychology of Aging (Eighth Edition).* Academic Press. Eds.: K. Warner Schaie, Sherry L. Willis. 165-185. ISBN 9780124114692, https://doi.org/10.1016/B978-0-12-411469-2.00009-1.

Witten, I. & Knudsen, E. (2005). Why Seeing Is Believing: Merging Auditory and Visual Worlds. *Neuron,* 48(3) 489-496. https://doi.org/10.1016/j.neuron.2005.10.020.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and Neural Correlates of Executive Functioning in Musicians and Non-Musicians. *PLOS ONE*, *9*(6), e99868. https://doi.org/10.1371/journal.pone.0099868

Appendices

Appendix 1 – The International Phonetical Alphabet (IPA) revised to 2018



By International Phonetic Association - http://linguistics.ucla.edu/people/keating/IPA/IPA_charts_2018.html (PNG, 600 ppi), CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=69255286

Appendix 2 – Links to videos of stimuli used

Female

Congruent /ba/

https://youtu.be/GVrmiYWWrRc.

Incongruent [Visual /ga/ - Auditory /ba/]

https://youtu.be/6_wcg7SjzAc.

Auditory only /ba/

https://youtu.be/O8IebVfYVOw.

Visual only /ba/

https://youtu.be/qELV8_IRxWg.

Male

Congruent /pa/

https://youtu.be/k0wByDQ8hH0.

Incongruent [Audio /pa/ Visual /ka/]

https://youtu.be/r23vTeVmQDA.

Auditory only /pa/

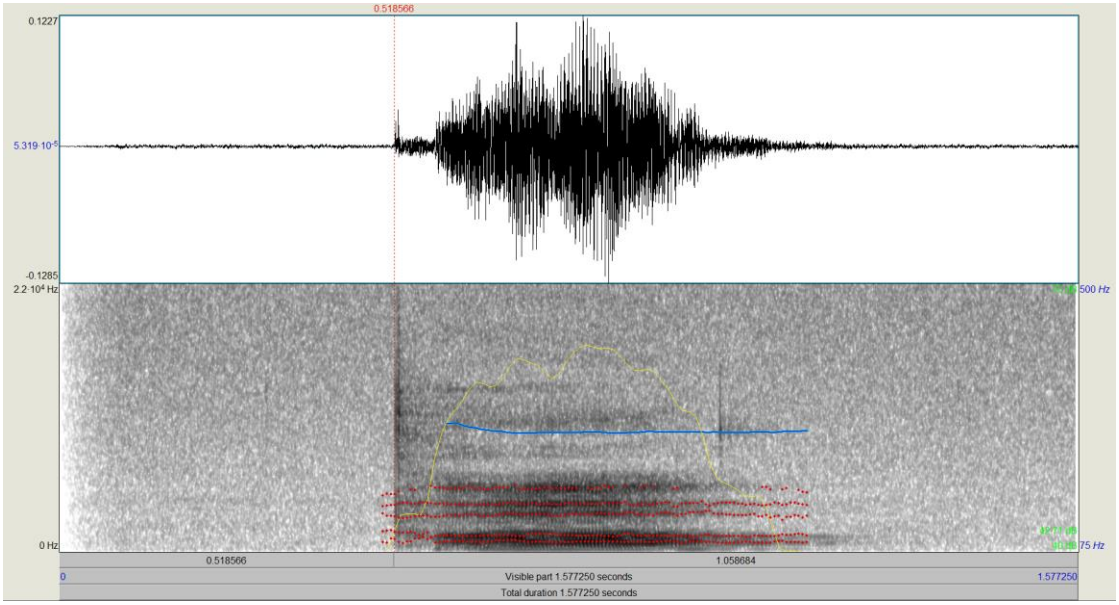
https://youtu.be/YOp5wr0W-xY.

Visual only /pa/

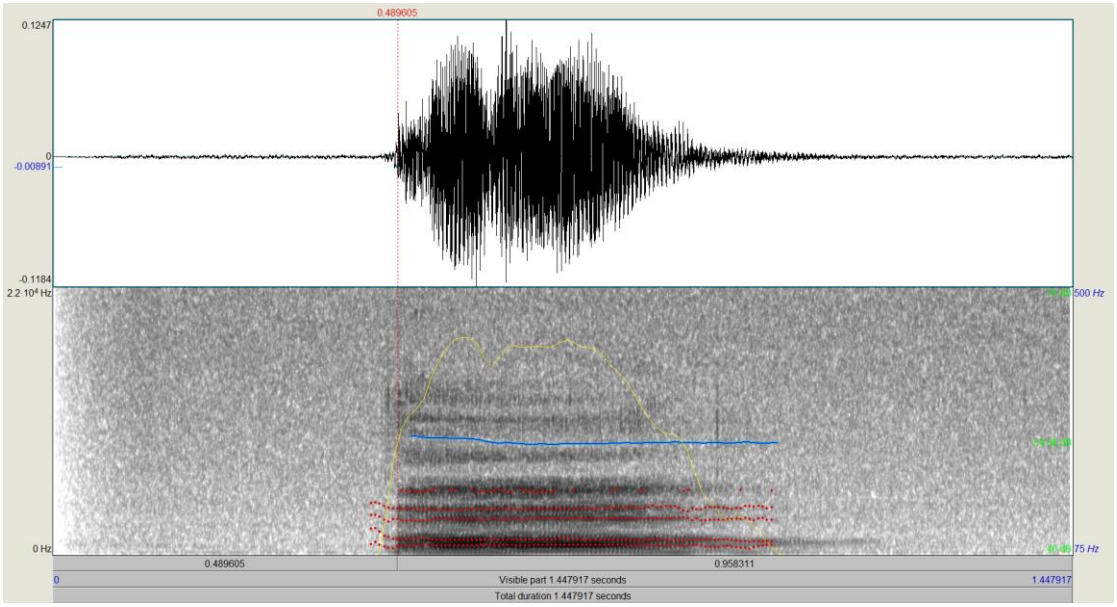
https://youtu.be/pXuW7HUQ9Zo.

Appendix 3 – Amplitude plots and spectrograms (female voice)

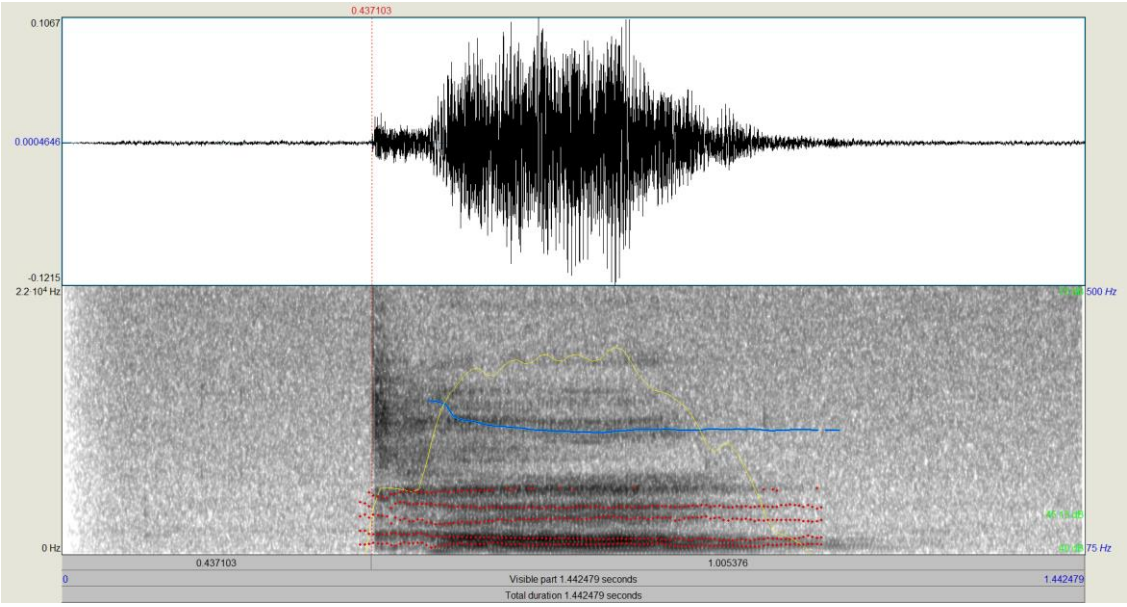
/pa/



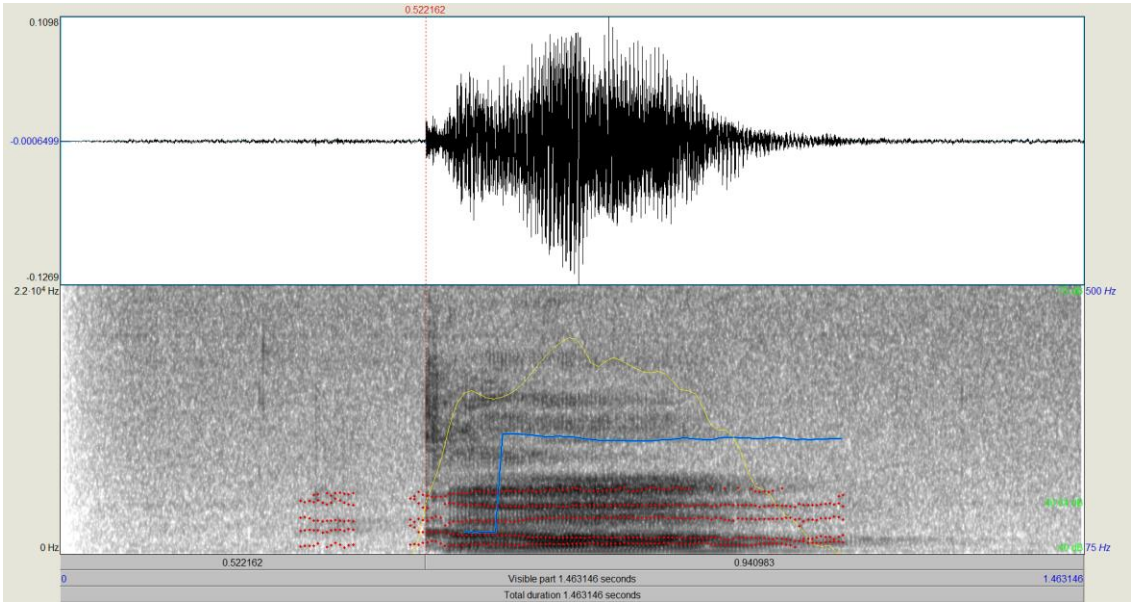
/ba/



/ka/

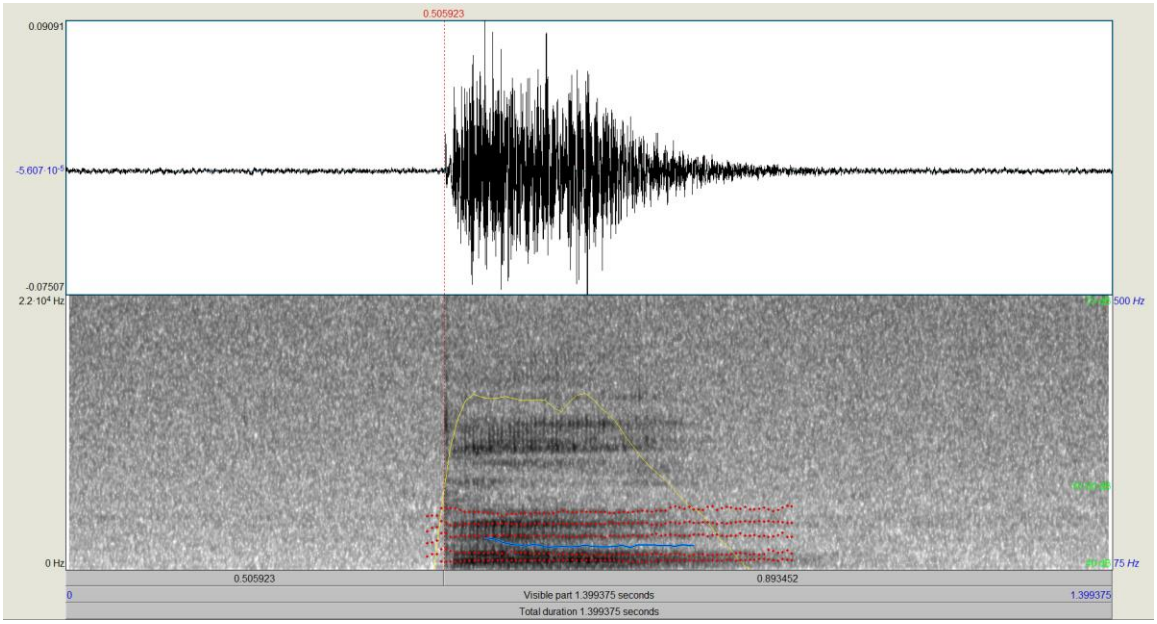

0.437103
0.1067
0.0004646
-0.1215
2.2·10⁴ Hz
500 Hz
0 Hz
75 Hz
0.437103
1.005376
0
Visible part 1.442479 seconds
1.442479
Total duration 1.442479 seconds

/ga/

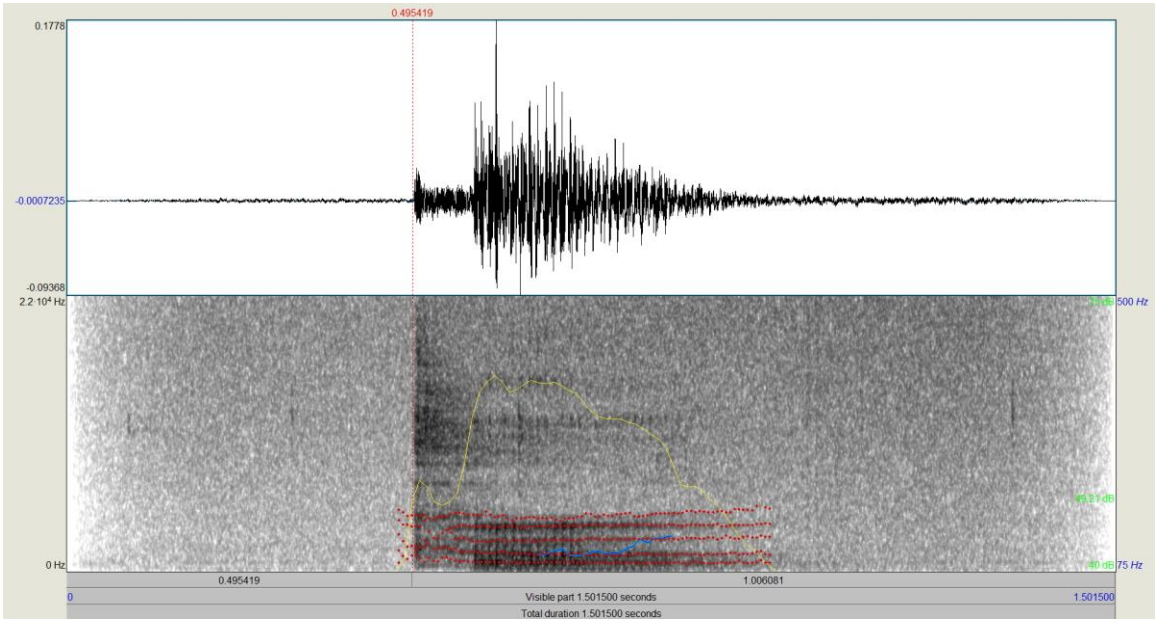

0.522162
0.1098
-0.0006499
-0.1269
2.2·10⁴ Hz
500 Hz
0 Hz
75 Hz
0.522162
0.940983
0
Visible part 1.463146 seconds
1.463146
Total duration 1.463146 seconds

Amplitude plots spectrograms (male voice)

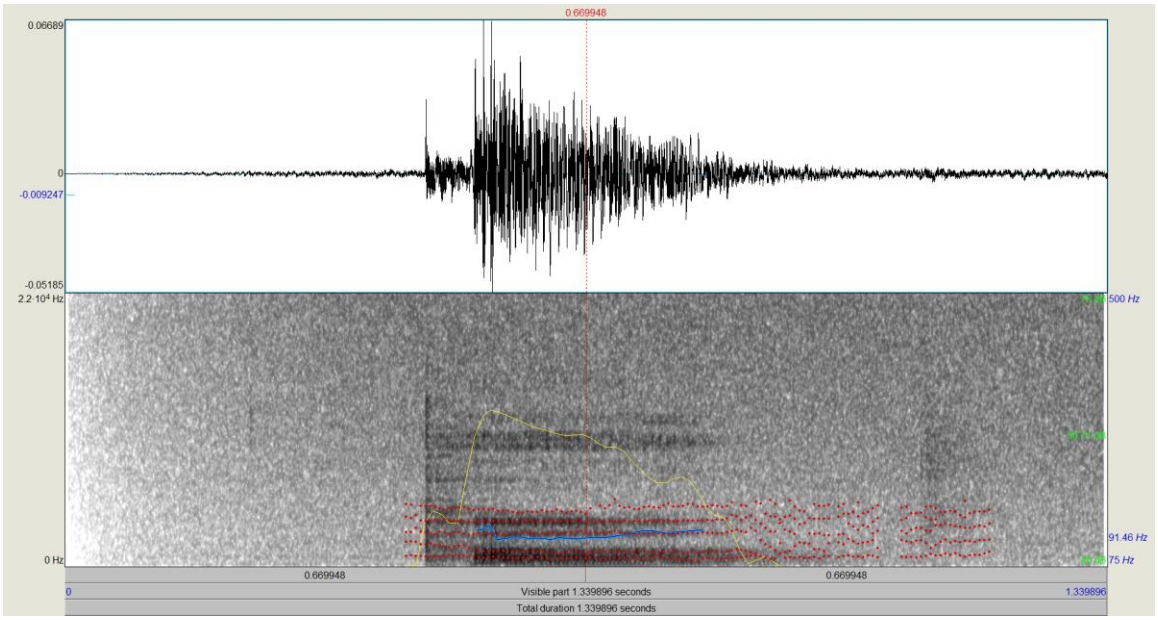
/ba/



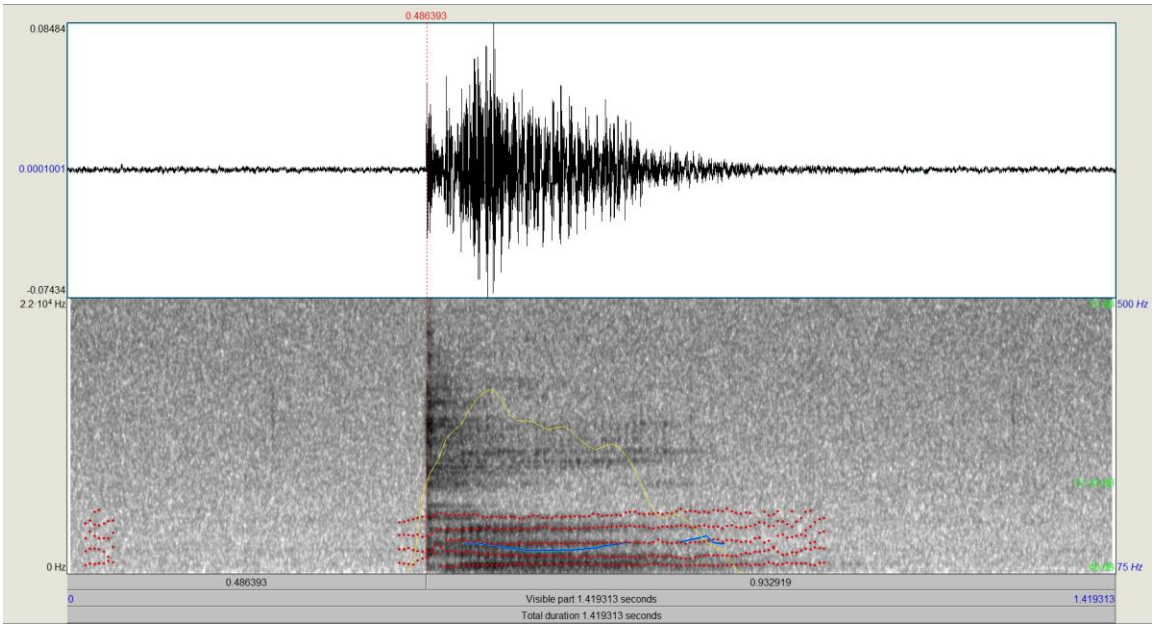
/ka/



/pa/


0.669948
0.06689
0
-0.009247
-0.05185
2.2·10⁴ Hz
500 Hz
91.46 Hz
0 Hz
75 Hz
0.669948
0.669948
0
Visible part 1.339896 seconds
1.339896
Total duration 1.339896 seconds

/ga/


0.486393
0.08484
0.0001001
-0.07434
2.2·10⁴ Hz
500 Hz
0 Hz
75 Hz
0.486393
0.932919
0
Visible part 1.419313 seconds
1.419313
Total duration 1.419313 seconds

Appendix 4 – Ethics forms

University of Victoria

Human Research Ethics Board

Application for Research Ethics Approval for Human Participant Research

The following application form is an institutional protocol based on the

Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans

Instructions:

1. Download this application and complete it on your computer. Hand written applications will not be accepted. You will receive a response from the HREB within 4-6 weeks.
2. Use the *Human Research Ethics Board Annotated Guidelines* to complete this application: http://www.uvic.ca/research/conduct/home/regapproval/humanethics/index.php.

   Note: This form is linked to the guidelines. Access links in blue text by hitting CTRL and clicking on the blue text.
3. Submit one (1) original and two (2) copies of this completed, signed application with all attachments to: Human Research Ethics, Michael Williams Building (MWB), Room B202, University of Victoria, PO Box 1700 STN CSC, Victoria BC V8W 2Y2 Canada
4. Do not staple the original copy (clips O.K.).
5. If you need assistance, contact the Human Research Ethics Office at (250) 472-4545 or ethics@uvic.ca
6. Please note that applications are screened and will not be entered into the review system if incomplete (e.g., missing required attachments, signatures, documents). You will be notified in this case.
7. Once approved, a Request for Annual Renewal must be completed annually for on-going projects for continuing Research Ethics approval.

A. Principal Investigator

If there is more than one Principal Investigator, provide their name(s) and contact information below in Section B, Other Investigator(s) & Research Team.

Last Name: **Tanaka** First Name: **James**

Department/Faculty: Psychology **UVic** Email: jtanaka@uvic.ca

Phone: (250) 721-7541 **Primary** Email:

Mailing Address (*if different from Department/Faculty*) including postal code:

Title/Position: (Must have a UVic appointment or be a registered UVic student)

☒ Faculty ☐ Undergraduate ☐ Ph.D. Student

☐ Staff ☐ Master's Student ☐ Post-Doctoral

☐ Adjunct or Sessional Faculty (Appointment start and end dates):_____

Students: Provide your Supervisor's information:

Name: Email:

Department/Faculty: Phone:

Graduate Students: Provide your Graduate Secretary's email address:

All PIs: Provide any additional contacts for email correspondence:

Name: **Juan Sebastian Vassallo** Email: **jvassallo@uvic.ca**

B. Project Information

Project Title: **Expert novice differences in speech perception.**

Anticipated Start Date for Recruitment / Data Collection: **SEPT 5** Anticipated End Date: **April 1 2018**

Geographic location(s) of study: University of Victoria

Participant recruitment/data collection location(s)/site(s): University of Victoria

Keywords: 1. McGurk 2. musicians 3. Bimodal perception 4.

Is this application connected/associated/linked to one that has been recently submitted? ☐ Yes ☒ No

If yes, provide further information:

All Current Investigator(s) and Research Team:

(*Include all current co-investigators, students, employees, volunteers, community organizations.)*

Contact Name	Role in Research Project	Institutional Affiliation	Email or Phone
Juan Vassallo	**Research assistant**	**UVic**	**jvassallo@uvic.ca**

For Faculty Only: Any Graduate Student Research Assistants who will use the data to fulfill UVic thesis/ dissertation/ academic requirements: Include all current Graduate Student Research Assistants

C. Multi-Jurisdictional Research

Does the proposed project require Research Ethics Board (REB) approval from another research ethics board(s)?

☐ Yes ☒ No

If yes, list the other research ethics board from which you or research team members have sought approval or will seek approval:

(Attach proof of having applied to other research ethics board(s). Please forward approvals upon receiving them. Be assured that UVic ethics approval may be granted prior to receipt of other research ethics board approvals.)

If you have answered "yes" above, please indicate your role in the multi-jurisdictional research project (Check all that apply):

☐ Recruiting participants

☐ Collecting data

☐ Analyzing data (with or without identifiers) collected by you and/or UVic research team members

☐ Analyzing data that *contains* identifiers: Data to be collected by non-UVic research team members as outlined in this application.

☐ Analyzing data that *does not* contain identifiers: Data to be collected by non-UVic research team members as outlined in this application.

☐ Dissemination of results via publications, reports, conferences, internet, etc.

☐ Other *(explain):*

D. Agreement and Signatures

For further information, on signature requirements, please see the Guidelines for Signatures.

Principal Investigator and Student Supervisor affirm that:

- *I have read this application and it is complete and accurate.*
- *The research will be conducted in accordance with the University of Victoria regulations, policies and procedures governing the ethical conduct of research involving human participants and all relevant sections of the TCPS 2.*
- *The conduct of the research will not commence until ethics approval has been granted.*

- *The researcher(s) will seek further HREB review if the research protocol is modified.*
- *Adequate supervision will be provided for students and/or staff.*

Principal Investigator

Student's Supervisor or co-Supervisor (for student applicants only)

_____ _____

Signature *Signature*

_____ _____

Print Name Print Name

_____ _____

Date Date

Chair, Director or Dean

(*To be signed by the person to whom the PI, or student's supervisor reports, and must not be the same person as the PI or student's supervisor. The Research Ethics Office cannot accept applications with duplicate signatures*)

I affirm that adequate research infrastructure is available for the conduct and completion of this research.

_____

Signature

_____

Print Name

_____

Date

E. Project Funding

Have you applied for funding for this project? ☐ Yes ☒ No If yes, please complete the following:

Source of Project Funding	**Funding Applied**	**Funding Approved**	**Project Title Used in Funding Application (or additional information)**
	☐ Yes ☐ No	☐ Yes ☐ No	
	☐ Yes ☐ No	☐ Yes ☐ No	
	☐ Yes ☐ No	☐ Yes ☐ No	
	☐ Yes ☐ No	☐ Yes ☐ No	

Will this project receive funding from the US *National Institutes of Health (NIH)*?

☐ Yes ☒ No

If yes, provide further information:

If you have applied for funding, have you submitted a funding application or contract notification to the UVic Office of Research Services?

☐ Yes ☒ No

F. Scholarly Review

What type of scholarly review has this research project undergone?

☐ External Peer Review (*e.g., granting agency*)

☐ Supervisory Committee or Supervisor—required for all student research projects

☒ None

☐ Other, please explain:

G. Other Approvals and Consultations

Do you require additional approvals or consultations from other agencies, community groups, local governments, etc.?

☐ Yes, attached ☐ Yes, will forward as received ☒ No

(Attach proof of having made request(s) for permission, or attach approval letter(s). Please forward approvals upon receiving them. Be assured that ethics approval may be granted prior to receipt of external approvals.)

If **Yes**, please check all that apply:

☐ **School District, Superintendent, Principal, Teacher.** Please list the school districts or schools:

☐ **BC Health Authorities and/or BC Universities.** Check all that apply:

☐ Island Health (VIHA)

☐ Interior Health (IH)

☐ Vancouver Coastal Health (VCH)

☐ Northern Health (NH)

☐ Fraser Health (FH)

☐ Simon Fraser University

☐ University of BC

☐ BC Cancer Agency

☐ Children's &Women's Hospital

☐ Providence Health Care

☐ University of Norther BC

If you are UVic faculty, student or staff and will be conducting research under the auspices of any of the institutions listed above, (involving staff, patients, health records, sites and/or recruitment through their sites, including recruitment via poster placement), your application may be reviewed under the BC Ethics Harmonization Initiative. (a single coordinated review with the other institution(s) listed). Harmonization also applies when members of your research team consist of faculty, staff and students from the BC institution(s) listed above. Please contact ethics@uvic.ca, 250-472-4545 if you have questions about a harmonized review.

Please explain:

☐ **Other regional government authority**, please explain:

☐ **Community Group (e.g., formal organization, informal collective),** please explain:

☐ **Other Research Ethics Board (REB) Approval**, please explain:

☐ **UVic Biosafety Committee Approval.** *Attach your Biosafety Approval, or your correspondence with the Biosafety Committee, to this application. Note that Research Ethics Approval is contingent on Biosafety Approval.*

☐ **Other Approval**, please explain:

H. Researcher(s) Qualifications

In light of your research methods, the nature of the research, and the characteristics of the participants, what training, qualifications, or personal experiences do you and/or your research team have (*e.g., research methods course, language proficiency, committee expertise, training on the equipment to be used)*?

All researchers and volunteers are trained on obtaining informed consent, giving experimental instructions, running computer tasks, and debriefing participants. All research assistants have completed the tutorial on human ethics found on the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans website: http://www.pre.ethics.gc.ca/eng/education/tutorial-didacticiel.

I. Research Involving Aboriginal Peoples of Canada (Including First Nations, Inuit and Métis)

The TCPS 2 (Chapter 9) highlights the importance of community engagement and respect for community customs, protocols, codes of research practice and knowledge when conducting research with Aboriginal peoples or communities. "Aboriginal peoples" includes First Nations, Inuit and Métis regardless of where they reside or whether or not their names appear on an official register. The nature and extent of community engagement should be determined jointly by the researcher and the relevant community or collective, taking into account the characteristics and protocols of the community and the nature of the research.

1. Conditions of the Research

1a. Will the research be conducted on (an) Aboriginal – First Nations, Inuit and Métis – lands, including reserves, Métis settlement, and lands governed under a self-government agreement or an Inuit or First Nations land claims agreement?

☒ No

☐ Yes, provide details:

1b. Do any of the criteria for participation include membership in an Aboriginal community, group of communities, or organization, including urban Aboriginal populations?

☒ No

☐ Yes, provide details:

1c. Does the research seek input from participants regarding a community's cultural heritage, artifacts, traditional knowledge or unique characteristics?

☐ Yes ☒ No

1d. Will Aboriginal identity or membership in an Aboriginal community be used as a variable for the purposes of analysis?

☐ Yes ☒ No

1e. Will the results of the research refer to Aboriginal communities, peoples, language, history or culture?

☐ Yes ☒ No

2. Community Engagement

2a. If you answered "yes" to questions a), b), c), d) or e), have you initiated or do you intend to initiate an engagement process with the Aboriginal collective, community or communities for this study?

☐ Yes ☐ No

2b. If you answered "yes" to question 2a, describe the process that you have followed or will follow with respect to community engagement. Include any documentation of consultations (*i.e. formal research agreement, letter of approval, email communications, etc*.) and the role or position of those consulted, including their names if appropriate:

3. No community consultation or engagement

If you answered "no" to question 2a, briefly describe why community engagement will not be sought and how you can conduct a study that respects Aboriginal communities and participants in the absence of community engagement.

J. International Research

4. Will this study be conducted in a country other than Canada?

☐ Yes ☒ No

If yes, describe how the laws, customs and regulations of the host country will be addressed (*consider research Visas, local Institutional Research Ethics Board requirements, etc.*):

K. Description of Research Project

5. Purpose and Rationale of Research

Briefly describe in non-technical language:

In this study we aim to research about audiovisual sensory integration in speech perception. Participant will be required to complete a survey about their musical background, and a computer task in which they are asked to recognize the syllable being shown in three conditions: audio of syllable plus video of person speaking a syllable, audio only and video only (see Appendix IV).

5a. The research objective(s) and question(s)

The research question for this experiment is whether expert musicians are better able to distinguish a syllable when shown with a video than novice participants..

5b. The importance and contributions of the research

It has been demonstrated that extensive musical training has powerful effects on many cognitive skills. Musicians are particularly skilled in auditory analysis and this might have long lasting effects on speech processing ability. This study aims to provide further evidence supporting differences between musicians and non-musicians in their ability to combine audio and visual information in speech perception.

5c. If applicable, provide background information or details that will enable the HREB to understand the context of the study when reviewing the application.

Not applicable

L. Recruitment

6. Recruitment and Selection of Participants

6a. Briefly describe the target population(s) for recruitment. Ensure that all participant groups are identified (*e.g., group 1 - teachers, group 2 - administrators, group 3 - parents*).

Group 1: Native English-speaking musicians:

No visual or auditory impairments. Instrumentalists / Singers with at least 8 years of continuous training in Western classical music on their principal instrument beginning at or before the age of 13, with formal private or group lessons within the past 5 years and currently played their instrument(s) for 5 or more hours a week.

Group 2: Native English-speaking non musicians

No visual or auditory impairments. Had no more than 3 years of formal music training on any combination of instruments throughout their lifetime nor had received formal instruction within the past 5 years.

6b. Why is each population or group of interest?

Group 1: Native English-speaking musicians will serve as our expert group.

Group 2: Native English-speaking non musicians will serve as our control group.

6c. What are the *salient* characteristics of the participants for your study? (*e.g., age, gender, race, ethnicity, class, position, etc*.)? List all inclusion and exclusion criteria you are using.

INCLUSION:

The multisensory nature of the experiment requires participants to have normal or corrected-to-normal vision (i.e., contacts or glasses are acceptable) and normal-to-corrected audition (i.e., hearing aids)

EXCLUSIONS:

Participants must be ages 18 years and up, native English-speakers. There are no other exclusion criteria

6d. What is the desired number of participants for each group?

***Group 1: Native English-speaking musicians* - A minimum of 12 participants.**

***Group 2: Native English-speaking non musicians* - A minimum of 12 participants**

6e. Provide a detailed description of your recruitment process. Explain:

i) List all source(s) for information used to contact potential participants (*e.g., personal contacts, listserves, publicly available contact information, etc.). Clarify which sources will be used for which participant groups:*

For Group 1 (*Native English-speaking musicians)* Expert musicians will be recruted using personal contacts, and from publicly available contact information (e.g., music department listserve, Facebook PAGE SPECIFICALLY CREATED FOR RECRUITING PARTICIPANTS TO THIS STUDY. INDIVIDUALS RECRUITED VIA FACEBOOK WILL BE INSTRUCTED TO CONTACT THE RESEARCHERS DIRECTLY INSTEAD OF RESPONDING PUBLICLY (I.E. VIA THE COMMENT AND REPLY FUNCTIONS).

For Group 2 (Native English-speaking non musicians) - The study will be advertised on the Psychology Research Participant System (SONA) and potential participants will be invited to contact the researchers through the Psychology Research Participation System if they are interested in participating. If researchers need to communicate with potential participants who express interest in participating, they will use the contact information provided through the SONA system.

ii) List all methods of recruitment (*e.g., in-person, by telephone, letter, snowball sampling, word-of-mouth, advertisement, etc.) If you will be using "snowball" sampling, clarify how this will proceed (i.e., will participants be asked to pass on your study information to other potential participants?). Clarify which methods will be used for which participant groups.*

The methods of recruitment will be: E-mail, In-person and word-of-mouth advertisement as "snowball" sampling: P*articipants will be asked to pass on your study information to other potential participants.* Also flyers will be posted in the School of Music of the University of Victoria, in the Conservatory of Music, and in the Victoria Symphony. UVIC SCHOOL OF MUSIC WILL BE CONTACTED TO OBTAIN PERMISSION TO USE THEIR LISTSERV. UVIC SCHOOL OF MUSIC, THE CONSERVATORY OF MUSIC, AND VICTORIA SYMPHONY WILL BE CONTACTED FOR PERMISSION TO POST FLYERS.

The experiment will be advertised online on the University of Victoria's Psychology Research Participation System (SONA) with a description of the experiment.

iii) If you will be using personal and/or private contact information to contact potential participants (as stated above), have the potential participants given permission for this, or will you use a neutral third party to

assist you with recruitment? *Note that this is not a concern when public and/or business contact information is used.*

Yes, by contacting the researchers on SONA, potential participants give researchers permission to contact them using the personal contact information that they provide on the SONA system.

iv) Who will recruit/contact participants (*e.g., researcher, assistant, third party, etc.) Clarify this for each participant group.*

The research assistant (Juan Vassallo) will manage the experiment posting on the Psychology Research Participation System and respond to requests to participate. THE RESEARCH ASSISTANT WILL SCREEN POTENTIAL GROUP 1 PARTICIPANTS (MUSICIANS) ACCORDING TO THE ELIGIBILITY CRITERIA CITED IN 6A.

v) List and explain any relationship between the members of the research team (including third party recruiters or sponsors/clients of the research) and the participant(s) (*e.g., acquaintances, colleagues*). Complete item 7 if there is potential for a power relationship or a *perceived* power relationship (*e.g., instructor-student, manager-employee, etc.*). If you have a close relationship with potential participants (*e.g., family member, friend, close colleague, etc.)* clarify here the safeguards that you will put in place to mitigate any potential pressure to participate.

It is expected that there will be no relationship between investigators and participants. However, some participants may coincidentally be acquaintances or classmates (both former and current) of the researchers.

vi) In chronological order (if possible) describe the steps in the recruitment process. *(Include how you will screen potential participants where applicable).* Consider where in the process permission of other bodies may be required.

A brief summary of the purpose, procedure, and length of the experiment will be posted on online recruitment website with a link for participation.

7. Power Relationships (Dual-Role and Power-Over)

If you are completing this section, please refer to the:

Guidelines For Ethics in Dual-Role Research for Teachers and Other Practitioners and the TCPS 2, Article 3.1 and Article 7.4.

Are you or any of your co-researchers in any way in a power relationship, including dual-roles, that could influence the voluntariness of a participant's consent? Could you or any of your co-researchers potentially be *perceived* to be in a power relationship by potential participants? *Examples of "power relationships" include teachers-students,*

therapists-clients, supervisors-employees and possibly researcher-relative or researcher-close friend where elements of trust or dependency could result in undue influence.

☐ Yes ☒ No ☐ Varies

If *yes* or *varies*, describe below:

i) The nature of the relationship:

ii) Why it is necessary to conduct research with participants over whom you have a power relationship:

iii) What safeguards (steps) will be taken to ensure voluntariness and minimize undue influence, coercion or potential harm:

iv) How will the power or dual-role relationship and associated safeguards be explained to potential participants:

Recruitment Materials Checklist:

Attach all documents referenced in this section *(check those that are appended):*

☐ Script(s) – in-person, telephone, 3rd party, e-mail, etc.

☒ Invitation to participate *(e.g., Psychology Research Participation System Posting)*

☐ Advertisement, poster, flyer

☐ None; please explain why *(e.g., consent form used as invitation/recruitment guide)*

M. Data Collection Methods

Data Collection

Use the following sections in ways best suited to explain your project. If you have more than one participant group, be sure to explain which participant group(s) will be involved in which activity/activities or method(s).

8a. Which of the following methods will be used to collect data? *Check all that apply.*

<table>
<tr><td>☐ Interviewing participants:

☐ in-person

☐ by telephone

☐ using web-based technology (explain):

☐ Conducting group interviews or discussions (including focus groups)</td><td>☐ Attach draft interview questions</td></tr>
<tr><td>☒ Administering a questionnaire or survey:

☒ In person ☐ by telephone

☐ mail back ☐ email

☐ web-based* (see below)

☐ Other, describe:

*If using a web program with a server located in the United States (e.g., SurveyMonkey), or if there are other reasons that the data will be stored in the US (e.g., use of US-based cloud technology, sharing data with US colleagues, etc.), you must inform participants that their responses may be accessed via the U.S. Freedom Act. Please add the following to the consent form(s):

“Please be advised that this research study includes data storage in the U.S.A. As such, there is a possibility that information about you that is gathered for this research study may be accessed without your knowledge or consent by the U.S. government in compliance with the U.S. Freedom Act. ”</td><td>☒ Attach questionnaire or survey:

☒ standardized (one with established reliability and validity)

☐ non-standardized (one that is un-tested, adapted or open-ended)</td></tr>
<tr><td colspan="2">☒ Administering a computerized task (describe in 8b or attach details)</td></tr>
<tr><td colspan="2">☐ Observing participants

In 8b, describe who and what will be observed. Include where observations will take place. If applicable, forward an observational data collection sheet for review.</td></tr>
<tr><td>☐ Recording of participants and data using:

☐ audio ☐ video ☐ photos or slides

☐ note taking ☐ flipcharts</td><td>☐ Images used for analysis</td></tr>
</table>

☐ data collection sheet *(attach)* ☐ other:	☐ Images used in disseminating results *(include release to use participant images in consent materials)*

☐ **Using human samples** (*e.g., saliva, urine, blood, hair*)

Attach your Biosafety Approval, or your correspondence with the Biosafety Committee, to this application. Note that Research Ethics Approval is contingent on Biosafety Approval.

☐ **Using specialized equipment/machines** *(e.g., ultrasound, EEG, prototypes etc.)* or other. *(e.g., testing instruments that are not surveys or questionnaires).* Please specify:

☐ **Using other testing equipment not captured under other categories.**

Please specify:

☐ **Collecting materials supplied by, or produced by, the participants** *(e.g., artifacts, paintings, drawings, photos, slides, art, journals, writings, etc.)*

Please specify:

☐ **Analyzing secondary data** or secondary use of data (Refers to information/data that was originally gathered for a purpose other than the proposed research and is now being considered for use in research (*e.g., patient or school records, personal writings, lesson plans, etc.).*

☐ Secondary data involving anonymized information (Information/data is stripped of identifiers by another researcher or institution **before** being shared with the applicant*).*

☐ Secondary data with identifying information (Data contains names and other information that can be linked to individuals,*(e.g., student report cards, employment records, meeting minutes, personal writings).*

In item 8b describe the source of the data, who the appropriate data steward is, and explain whether (and how) consent was or will be obtained from the individuals for use of their data.

☐ **Other:**

Please specify:

8b. Provide a sequential description of the procedures/methods to be used in your research study. Be sure to provide details for all methods checked in section 8a. Clarify which procedures/methods will be used for each participant group. Indicate which methods, if any, will be conducted in a group setting. *List all of the research instruments and interview/focus group questions, and append copies (if possible) or detailed descriptions of all instruments. If not yet finalized, provide drafts or sample items/questions.*

Prior to testing, the procedure will be thoroughly explained and informed written consent will be obtained from the participant. Participants will be able to ask any questions they may have in regards to the experiment.

PRIOR TO THE COMPUTER TESTING, ALL PARTICIPANTS (GROUPS 1 AND 2) WILL BE ASKED TO FILL OUT A QUESTIONNAIRE ABOUT THEIR MUSICAL BACKGROUND. THE QUESTIONNAIRE WILL BE ADMINISTERED ON THE COMPUTER USING THE UNIVERSITY OF VICTORIA'S SURVEYMONKEY PLATFORM. VERBAL INSTRUCTIONS WILL BE PROVIDED BY AN EXPERIMENTER, AS WELL AS WRITTEN INSTRUCTIONS ON THE QUESTIONNAIRE.

The experiment will be conducted using a laboratory computer. All instructions will be provided verbally by an experimenter, as well as in writing on the computer monitor.

In the experiment, participants will be shown videos with faces pronouncing syllables (see Appendix 4 for stimuli examples) on the computer screen and will be asked to recognize the syllable being uttered and to input responses pressing buttons on a keyboard.

8c. Where will participation take place for each data collection method/procedure? *Provide specific location*, (*e.g., UVic classroom, private residence, participant's workplace). Clarify the locations for each participant group and/or each data collection method*.

ɪrticipation will take place in the University of Victoria's Visual Cognition Lab, located in Cornett A081.

8d. For each method, and in total, how much time will be required of participants? *Clarify this for each participant group, each data collection method, and any other research related activities.*

Survey = 10 minutes, task = 30 minutes, debrief = 5 minutes; total = 45 minutes.

8e. Will participation take place during participants' office/work hours or instructional time?

☒ No ☐ Yes. Indicate whether permission is required *(e.g., from workplace supervisor, school principal, etc.)* and how this will be obtained:

Data Collection Methods Checklist:

Attach all documents referenced in this section *(check those that are appended. Where draft versions are appended please ensure that final versions are submitted when available. If final versions differ significantly after you have obtained Research Ethics approval, you will need to submit a Request for Modification:*

☐ Standardized Instrument(s)

☒ Survey(s), Questionnaire(s)

☐ Interview and/or Focus Group Questions

☐ Observation Protocols

☐ Other:

N. Possible Benefits, Inconveniences, and Risks of Harm to Participants

9. Benefits

Identify any potential or known benefits associated with participation and explain below. *Keep in mind that the anticipated benefits should outweigh any potential risks.*

☒ To the participant ☐ To society ☐ To the state of knowledge

By participating in this research, participants will add to their knowledge of how experimental psychologists conduct research and learn about issues pertaining to human perception. The debriefing will include the theoretical importance of the research, as well as examples of how the results will benefit and/or clarify our understanding of the issues under investigation. Therefore, participants will be given an opportunity to learn about current topics in experimental psychology.

10. Inconveniences

Identify and describe any known or potential inconveniences to participants: *Consider all potential inconveniences, including total time devoted to the research.*

Participants are required to devote approximately 45 minutes of their time to this experiment. Eye strain or fatigue may occur but are unlikely, since the computer task is about 30 minutes and is conducted in a brightly lit room, and breaks will be provided.

11. Level of Risk

The TCPS 2 definition of "minimal risk research" is as follows:

> *"Research in which the probability and magnitude of possible harms implied by participation in the research is no greater than those encountered by the participant in those aspects of their everyday life that relate to the research."*

Based on this definition, do you believe your research qualifies as "minimal risk research"?

☒ Yes it is minimal risk. ☐ No, it is not minimal risk.

Explain your answer with reference to the risks of the study and the vulnerability of the participants:

There is no risk for participating in this experiment, apart from a minimal possibility of visual and / or auditory exhaustion.

12. Estimate of Risks of Harm

Consider the inherent foreseeable risks associated with your research protocol and complete the table below by putting an X in the appropriate boxes. Be sure to take into account the vulnerability of your target population(s) if applicable:

Potential Risks of Harm	**Very unlikely**	**Possibly**	**Likely**
i) Emotional or psychological discomfort, such as feeling demeaned or embarrassed due to the research	☒	☐	☐
ii) Fatigue or stress	☐	☒	☐

iii) Social risks, such as stigmatization, loss of status, privacy and/or reputation	☒	☐	☐
iv) Physical risks such as falls	☒	☐	☐
v) Economic risk (e.g., job security, salary loss, etc.)	☒	☐	☐
vi) Risk of incidental findings (*See Article 3.4 of the TCPS 2 for more information*)	☒	☐	☐
vii) Other risks:	☒	☐	☐

13. Possible Risks of Harm

If you indicated in Item 12 (i) to (vii) that any risks of harm are *possible* or *likely*, please explain below:

13a. What are the risks? *(i.e., elaborate on risks you have identified above)*

Participants may experience fatigue or eye strain due to the task being administered on a computer.

13b. What will you do to try to minimize, mitigate, or prevent the risks?

Fatigue and eyestrain:

Participants will also be encouraged to take breaks as necessary throughout the computerized task to minimize strain on their eyes.

13c. How will you respond if the harm occurs? *(i.e., what is your plan?)*

Participants will be encouraged to withdraw from the experiment if they experience any discomfort and they will be told that they can do so without any penalty or loss of compensation. Any participant who expresses discomfort will be given information about UVic counselling services.

13d. If you have indicated that there is a risk of Incidental Findings (vi) please outline your proposed protocol for information and/or action.

Not applicable.

13e. If one or more of your participant groups could be considered vulnerable please describe any specific considerations you have built into the protocol to address this.

Not applicable.

14. Risk to Researcher(s)

14a. Does this research study pose any risks to the researchers, assistants and data collectors?

No

14b. If there are any risks, explain the nature of the risks, how they will be minimized, and how you will respond if they occur.

15. Deception

Will participants be fully informed of everything that will be required of them prior to the start of the research session?

☒ Yes ☐ No *(If no, complete the Request to Use Deception form on the ORS website)*

O. Incentives, Reimbursement and Compensation

16a. Is there any incentive, monetary or otherwise, being offered for participation in the research *(e.g., gifts, honorarium, course credits, etc.)*

☒ Yes ☐ No

If yes, explain the nature of the incentive(s) and why you consider it necessary. *Also consider whether the amount or nature of the incentive could be considered a form of undue inducement or affect the voluntariness of consent. Clarify which participant groups will be provided with which incentives*.

Participants recruited from the Psychology Research Participation System will receive course credit as compensation. This is an adequate reimbursement for the time and effort required to participate, but does not constitute undue enticement. The participation of musicians is strictly volunteer. Detailed information about music and speech perception will be offered, and also the experience of being part of a Psychology experiment.

16b. Is there any reimbursement or compensation for participating in the research *(e.g., for transportation, parking, childcare, etc.)*

☐ Yes ☒ No

If yes, explain the nature of reimbursement or compensation and why you consider it necessary. *Also consider whether the amount of reimbursement or compensation could be considered a form of undue inducement or affect the voluntariness of consent. Clarify which participant groups will be provided with which kind of reimbursement or compensation.*

16c. Explain what will happen to the incentives, reimbursement or compensation if participants withdraw during data collection or any time thereafter *(e.g., compensation will be pro-rated, full compensation will be given, etc.)*

Full compensation will be given to participants who show up to participate, yet withdraw during or after data collection.

P. Free and Informed Consent

Consent encompasses a process that begins with initial contact and continues through to the end of the research process. Consult Article 3.2 of the TCPS 2 and Appendix V of the Guidelines for further information.

17. Participant's Capacity (Competence) to Provide Free and Informed Consent

Capacity refers to the ability of prospective or actual participants to understand relevant information presented about a research project, and to appreciate the potential consequences of their decision to participate or not participate. See the TCPS 2, Chapter 3, section C, for further information.

Identify your potential participants: *(Check all that apply.)*

Competent	**Non-Competent**
☒ Competent adults ☐ A protected or vulnerable population (*e.g., inmates, patients*)	☐ Non-competent adults: ☐ Consent of family/authorized representative will be obtained ☐ Assent of the participant will be obtained (note that assent of the participant is always required)
☐ Competent youth aged 13 to 18:	☐ Non-competent youth: ☐ Consent of parent/guardian ☐ Assent of the youth will be obtained (note that assent of the participant is always required)

☐ Consent of youth will be obtained and parental/guardian consent is required, *due to institutional requirements (such as school districts) or due to the nature of the research (e.g., risks, etc.)* ☐ Consent of youth will be obtained, parents/guardians will be informed ☐ Consent of youth will be obtained, parents/guardians will *NOT* be informed ☐ Other, explain:	
☐ Competent children under 13 (*who are able to provide fully informed consent)*: ☐ Consent of child will be obtained and consent of parent/guardian will be obtained ☐ Other, explain:	☐ Non-competent children *(young children and/or children with limited abilities to provide fully informed consent)*: ☐ Consent of parent/guardian ☐ Assent of the child will be obtained (note that assent of the participant is always required)

18. Means of Obtaining and Documenting Consent and/or Assent:

Check all that apply, consider all of your participant groups, attach copies of relevant materials, complete item 19:

☒ **Signed** consent (*Attach consent form(s) - see template available)*

☐ **Verbal** consent (*Attach verbal consent script(s) - see template available*.)

Explain in 19 why written consent is not appropriate and how verbal consent will be documented.

☐ Letter of Information for **Implied** consent (*e.g., anonymous, mail back or web-based survey. Attach information letter, see template)*

☐ **Signed** or **Verbal assent** for non-competent participants *(Attach assent form(s), or verbal assent script(s)).*

Explain how verbal assent will be documented in 19.

☐ **Other** means. **Explain** in 19 and provide justification.

☐ Consent **will not be obtained**. See TCPS 2 Articles 3.5 and 3.7. **Explain** in 19**.**

☐ **Signed** consent from the parents/guardians for youth/child participants ***(****Attach consent form(s)).*

Explain how parents/guardians will provide informed consent for child/youth participants in 19.

☐ **Information letters** for the parents/guardians of youth/child participants *(Attach information letter(s)). If consent will not be obtained from parents/guardians and the parents/guardians will not be informed, explain why not in 19.*

19. Informed Consent

Describe the exact steps (chronological order) that you will follow in the process of explaining, obtaining, and documenting informed consent. Ensure that consent procedures for all participant groups are identified (e.g., group 1 - teachers, group 2 – parents, group 3 – students). Be sure to indicate when participants will first be provided with the consent materials *(e.g., prior to first meeting with the researcher?).* If consent will not be obtained, explain why not with reference to the TCPS 2 Articles 3.5 and 3.7.

A brief description of the experimental procedures will be given to the participants verbally at the start of the session (although they may read the instructions as well). Participants will then read and sign a written consent form.

20. Ongoing Consent

Article 3.3 of the TCPS 2 states that consent shall be maintained throughout the research project. Complete this section if the research involves interacting with participants over multiple occasions (including review of transcripts, etc.), has multiple data collection activities, and/ or occurs over an extended period of time.

20a. Will your research occur over multiple occasions or an extended period of time *(including review of transcripts)*?

☐ Yes ☒ No

20b. If yes, describe how you will obtain and document ongoing consent. If consent procedures differ for each group or activity, please clarify each group or activity that you are referring to.

21. Participant's Right to Withdraw

Article 3.1 of the TCPS2 states that participants have the right to withdraw at any time and can withdraw their data and human biological materials.

Describe what participants will be told about their right to withdraw from the research at any time (*i.e., who to contact and how*). If compensation is involved, explain what participants will be told about compensation if they withdraw. *If you have different participant groups and/or different data collection methods, clarify the different procedures for withdrawing as necessary.*

Participants will be told that they can withdraw from the experiment at any time. Upon withdrawing they will be given full course credit.

22. What will happen to a person's data they withdraw part way through the study or after the data have been collected/submitted? If applicable, include information about visual data such as photos or videos. *If you have different participant groups and/or different data collection methods, clarify the different procedures for withdrawing as necessary. Ensure this information is included in the consent documents.*

☐ Participant will be asked if they agree to the use of their data. Describe how this agreement will be documented:

☒ It will not be used in the analysis and will be destroyed.

☐ It is logistically impossible to remove individual participant data (*e.g., anonymously submitted data*).

☐ When linked to group data (*e.g., focus group discussions*), it will be used in summarized form with no identifying information.

Free and Informed Consent Checklist:

Attach all documents referenced in this section *(check those that are appended):*

☒ Consent and Assent Form(s) – Include forms for all participant groups and data gathering methods

☐ Letter(s) of Information for Implied Consent

☐ Verbal Consent and Assent Scripts

Q. Anonymity and Confidentiality

23. Anonymity

Anonymity means that no one, including the principal investigator, is able to associate responses or other data with individual participants.

23a. Will the participants be anonymous in the data gathering phase of research?

☒ Yes ☐ No

23b. Will the participants be anonymous in the dissemination of results *(be sure to consider use of video, photos)*?

☒ Yes

☐ Maybe. Explain below.

☐ No. If anonymity will not be protected and you plan to identify all participants with their data, provide the rationale below.

24. Confidentiality

Confidentiality means the protection of the person's identity (anonymity) and the protection, access, control and security of his or her data and personal information during the recruitment, data collection, reporting of findings, dissemination of data (if relevant) and after the study is completed (e.g., storage). The ethical duty of confidentiality refers to the obligation of an individual or organization to safeguard entrusted information. The ethical duty of confidentiality includes obligations to protect information from unauthorized access, use, disclosure, modification, loss or theft.

24a. Are there any limits to protecting the confidentiality of participants?

☒ No, confidentiality of participants and their data will be completely protected

☐ Yes, there are some limits to the researcher's ability to protect the confidentiality of participants (*Check relevant boxes below.)*

☐ Limits due to the nature of group activities *(e.g., focus groups):* The researcher cannot guarantee confidentiality

☐ Limits due to context: The nature or size of the sample from which participants are drawn makes it possible to identify individual participants *(e.g., school principals in a small town, position within an organization)*

☐ Limits due to selection: The procedures for recruiting or selecting participants may compromise the confidentiality of participants *(e.g., participants are identified or referred to the study by a person outside the research team)*

☐ Limits due to legal requirements for reporting *(e.g., legal or professional)*

☐ Limits due to local legislation such as the U.S.A. Freedom Act *(e.g., when there will be data storage in the United States).* When using USA based data instruments and data storage systems researchers are responsible for determining if this applies.

☐ Other:

24b. If confidentiality will be protected, describe the procedures to be used to ensure the anonymity of participants and for preserving the confidentiality of their data *(e.g., pseudonyms, changing identifying information and features, coding sheet, etc.) If you will use different procedures for different participant groups and/or different data methods be sure to clarify each procedure*.

Each participant will be assigned a number code to identify their data while preserving their anonymity. The data will be stored in locked offices on password protected computers in the Cornett building. Only members of the experiment's lab group will have access to them.

24c. If there are limits to confidentiality indicated in section 24a. above, explain what the limits are and how you will address them with the participants. *If there are different procedures for different participant groups and/or different data collection methods, be sure to clarify each procedure*.

R. Use and Disposal of Data

25. Use(s) of Data

25a. What use(s) will be made of all types of data collected *(field notes, photos, videos, audiotapes, transcripts, etc.)*?

Data will be analyzed using conventional methods. Results may be submitted for publication in academic journals and/or presented at scientific meetings.

25b. Will your research data be analyzed, now or in future, by yourself for purposes other than this research project?

☐ Yes ☐ No ☒ Possibly

25c. If yes or possibly, indicate what purposes you plan for this data and how will you obtain consent for future data analysis from the participants *(e.g., request future use in current consent form)*?

Data may be used for re-analysis. Future use will be requested in the current consent form.

25d. Will your research data be analyzed, now or in future, by other persons for purposes other than explained in this application?

☐ Yes ☐ No ☒ Possibly

25e. If yes or possibly:

i) Indicate whether the data will contain identifiers when it is provided to the other researchers or whether it will be fully anonymous *(note that "fully anonymous" means that there is no identifying information, links, keys, or codes that allow the data to be re-identified).*

The data will be fully anonymous when provided with no identifying information, links, keys, or codes that allow the data to be re-identified.

ii) How will you obtain consent from the participants for future data analysis by other researchers? *(If the data will be transferred in fully anonymous form, this request for future use can be made in the current consent form. If the data will contain identifiers or links/keys/codes for re-identification, consider requesting permission to contact the participants in the future, to obtain consent for the use of the data at that time).*

Future data analysis by other researchers will be requested in the current consent form.

26. Commercial Purposes

26a. Do you anticipate that this research will be used for a commercial purpose?

☐ Yes ☒ No

26b. If yes, explain how the data will be used for a commercial purpose:

26c. If yes, indicate if and how participants will benefit from commercialization.

27. Maintenance and Disposal of Data

Describe your plans for protecting data during the project, and for preserving, archiving, or destroying all the types of data associated with the research (*e.g., paper records, audio or visual recordings, electronic recordings, coded data*) after the research is completed:

27a. means of storing and securing data (*e.g., encryption, password protected computer files, locked cabinet, separation of key codes from raw data etc.*):

Paper consent forms will be stored in locked filling cabinets.

To protect participant confidentiality, participant data will be identified using key codes and the key codes will be stored in a locked filling cabinet separate from the raw data. Computer data will be password-protected.

27b. location of storing data *(include location of data-storage servers if using web-based technology)*:

Paper consent forms will be stored in locked filling cabinets in Cornett A081 (Different Minds Lab)

Data collected from the computerized perception task will be stored on local drives of password protected computers in Cornett A081.

Data collected from the questionnaire will be stored on secure servers in Canada (SurveyMonkey, through the University of Victoria license for researchers and staff).

27c. duration of data storage *(if data will be kept indefinitely, explain why this is necessary and state whether the data will contain identifiers or links to identifiers)*:

Electronic data will be stored indefinitely. Consent forms will be kept for five years.

27d. methods of destroying or archiving data. If archiving data, please describe measures to secure or protect the data. If the archiving will involve a third party *(e.g., library, community agency, Aboriginal band, etc.)* please provide details:

Data will be destroyed by completely deleting computer files and shredding paper data.

28. Dissemination

How do you anticipate disseminating the research results? (*Check all that apply)*

☒ Thesis/Dissertation/Class presentation

☒ Presentations at scholarly meetings ☒ Published article, chapter or book

☐ Internet (*Students: Most UVic Theses are posted on "UVicSpace" and can be accessed by the public*)

☐ Media (*e.g., newspaper, radio, TV*)

☐ Directly to participants and/or groups involved. Indicate how: *(e.g., report, executive summary, newsletter, information session):*

☐ Other, explain:

S. Conflict of Interest

29a. Apart from a declared dual-role relationship (Section K, item 7), are you or any of the research team members in a perceived, actual or potential conflict of interest regarding this research project (*e.g., partners in research, private interests in companies or other entities*)?

☐ Yes ☒ No

29b. If yes, please provide details of the conflict and how you propose to manage it:

Attachments*

*Ensure that all applicable attachments are included with all copies of your application.

Information for Submission

- Applications may be printed and submitted double-sided
- Do **not** staple the original application with original signatures (clips O.K.)
- The two photocopies may be individually stapled or clipped
- Do **not** staple or clip the individual appendices

Title and label attachments as Appendix 1, 2, 3 etc. and attach the following documents (check those that are appended):

Section I - Recruitment Materials:

☐ Script(s) – in-person, telephone, 3rd party, e-mail, etc.

☒ Invitation to participate

☐ Advertisement, Poster, Flyer

Section J - Data Collection Methods:

☐ Standardized Instrument(s)

☒ Survey(s), Questionnaire(s)

☐ Interview and/or Focus Group Questions

☐ Observation Protocols

☐ Other:

Section M - Free and Informed Consent:

☒ Consent Form(s) – Include forms for all participant groups and data gathering methods

☐ Assent Form(s)

☐ Letter(s) of Information for Implied Consent

☐ Verbal Consent Script

☐ Approval from external organizations (or proof of having made a request for permission)

☐ Permission to gain access to confidential documents or materials

☐ Request to Use Deception form

☐ Biosafety Committee Approval

☐ Other, please describe:

Invitation to Participate on SONA

Study Name

Expert novice differences in speech perception

Abstract

It has been demonstrated that extensive musical training has powerful effects on many cerebral domains. Musicians are obviously particularly skilled in auditory analysis and this might have long lasting effects on speech processing ability. This study aims to provide further evidence supporting differences between musicians and non-musicians in the functionality related to audiovisual integration, particularly in speech perception.

Description

Participant will be required to complete a survey about their musical background, and a computer task in which they are asked to recognize the syllable being uttered in different videos.

Participation in this study is entirely voluntary. If at any point you wish to withdraw from the experiment, before or after agreeing to participate, there will be no penalty and there will be no adverse effects on your ability to participate in future studies.

All data will be stored securely in a filing cabinet and a password protected computer in the Different Minds Lab at UVic. Only the principal researcher and lab assistants will have access to it. The data will be coded by a numeric system and analyzed as group data. Your privacy will be protected in any scientific publication or presentation resulting from this study and individual participants will not be identified.

Benefits of participating in this study include the development of new knowledge about bimodal speech perception in musicians and non-musicians. Risks include the possibility of fatigue or eyestrain due to the computerized task and the minor possibility of discomfort associated with answering questions about your romantic attachment style.

A verbal debriefing will be provided once the study is over. Please feel free to contact Juan Vassallo at jvassallo@uvic.ca if you have any further questions. Meet at: Cornett A081

Eligibility Requirements

Participants must be native English-speakers and have normal or corrected to normal vision and hearing.

Sign-Up Restrictions

None

Duration

30 minutes

Credits

1 Credit

Researchers

Juan S. Vassallo

Email: **jvassallo@uvic.ca**

Consent Form

About this study:

This study aims to provide further evidence supporting differences between musicians and non-musicians in the functionality related to audiovisual integration, particularly in speech perception. If you choose to participate in this study, you will be required to complete a survey about your musical training, and a computer task in which you will be asked to recognize the syllable being uttered in different videos. The experiment will take approximately 45 minutes to complete, with breaks.

Risks and benefits:

The risks involved in the study are minimal. Some observers may experience fatigue or eyestrain as a result of computer viewing, so it is recommended that you take advantage of scheduled breaks or rest when needed. There is a minor risk of discomfort associated with answering personal questions about romantic attachment. However, you may find participation is beneficial and interesting as it may make you think about your own perceptual processes in different ways and provide you with insight into how psychological research is conducted.

Compensation:

You are being invited to participate in this study either because you have signed up for course credit in a Psychology class, or because you have contacted the laboratory to serve as a volunteer. This form of compensation must not be coercive. It is unethical to provide undue compensation or inducements to research participants. If you would not participate if the compensation was not offered, then you should decline.

Voluntary participation:

Your participation is voluntary and you have the right to withdraw at any time without consequence, loss of compensation and without having to provide an explanation. If you withdraw from the study, your data will be deleted from the computer. You may ask questions of the experimenter at any time so that you can fully understand the procedures we will be following. However, answers that may influence the experimental outcome may be deferred until the end of the experiment.

Confidentiality and anonymity:

Confidentiality and anonymity for participants is a primary concern for us, and will be protected by storing consent forms, as well as any other related materials, in a locked filing cabinet in the laboratory of Dr. James Tanaka (Cornett A081). Consent forms will be destroyed after five years to protect anonymity. The data will not be disseminated publicly, except in summary form through professional journals or books or scholarly meetings. Under no circumstances will individual subjects be identified. Data may be used in future research or shared with collaborating researchers, but will remain fully anonymous, with no identifying information.

Data will be anonymized and stored electronically by the researchers indefinitely. You may verify the ethical approval of this study, or raise any concerns you might have, by contacting the Human Research Ethics Office at the University of Victoria (250-472-4545 or ethics@uvic.ca).

Your signature below indicates that you understand the above conditions of participation in this study and that you have had the opportunity to have your questions answered by the experimenter. A COPY OF THIS CONSENT WILL BE LEFT WITH YOU, AND A COPY WILL BE TAKEN BY THE RESEARCHER.

_____	_____	_____
Name of Participant	*Signature*	*Date*

Appendix III: Survey about musical background and training.

This survey allows us to have more information on your musical background and training. Please answer the following questions:

1. How many years have you practiced music?

2. Can you correctly label or produce pitches without a reference pitch?

3. What is your main instrument? (if you are a singer write ‘voice’)

4. What age did you start practicing? (informal / not studying)

5. What age did you start studying / taking lessons / formal training?