2015-03-01

# Brain Mapping of the Latency Epochs in a McGurk Effect Paradigm in Music Performance and Visual Arts Majors

Lauren Donelle Nordstrom
*Brigham Young University - Provo*

# Brain Mapping of the Latency Epochs in a McGurk Effect Paradigm in Music Performance and Visual Arts Majors

Lauren Donelle Nordstrom

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

David L. McPherson, Chair
Richard W. Harris
Erin D. Bigler

Department of Communication Disorders

Brigham Young University

March 2015

ABSTRACT

Brain Mapping of the Latency Epochs in a McGurk Effect Paradigm in
Music Performance and Visual Arts Majors

Lauren Donelle Nordstrom
Department of Communication Disorders, BYU
Master of Science


The McGurk effect is an illusion that occurs when an auditory /ba/ is combined with a visual /ga/. The two stimuli fuse together which leads to the perception of /da/, a sound in between /ba/ and /ga/. The purpose of this study was to determine whether music performance and visual arts majors process mismatched auditory and visual stimuli, like the McGurk effect, differently. Nine syllable pairs were presented to 10 native English speakers (5 music performance majors and 5 visual arts majors between the ages of 18 and 28 years) in a four-forced-choice response paradigm. Data from event-related potentials were recorded for each participant. Results demonstrate that there are differences in the electrophysiological responses to viewing the mismatched syllable pairs. The /ga/ phoneme in the music performance group produced more differences while the /da/ phoneme produced more differences in the visual arts group. The McGurk effect is processed differently in the music performance majors and the visual arts majors; processing begins in the earliest latency epoch in the visual arts group but in the late latency epoch in the music performance group. These results imply that the music performance group has a more complex decoding system than the visual arts group. It also may suggest that the visual arts group is better able to integrate the visual and auditory information to resolve the conflict when mismatched signals are presented.

Keywords: auditory perception, brain mapping, dipole localization, electroencephalography, event-related potentials, visual perception

ACKNOWLEDGMENTS

I would like to express my gratitude to my thesis chair, Dr. McPherson, for mentoring me through the process of completing a master's thesis. Without his expertise, guidance, advice, and constant encouragement, I could not have completed this project. I would also like to thank my committee members, Dr. Harris and Dr. Bigler for their suggestions and advice. I would like to thank Mark McPherson for creating the program and electronic computer interface that was used to present my stimuli, thus enabling me to collect my data. This study could not have moved forward without this crucial component. In addition, this project would not have been possible without the sacrifice of time from my participants. In spite of their busy schedules, these students made time to help a fellow student. Lastly, I would like to thank my family and friends who never wavered in their support and encouragement. They believed in me and did not doubt my ability to reach my goal.

TABLE OF CONTENTS

# LIST OF TABLES

Table Page

# LIST OF FIGURES

Figure Page

# LIST OF APPENDICES

Appendix Page

# DESCRIPTION OF THESIS STRUCTURE

This thesis is part of a larger research project, and portions of this thesis may be published as part of articles listing the thesis author as a co-author. The body of this thesis is written as a manuscript suitable for submission to a peer-reviewed journal in speech-language pathology. An annotated bibliography is presented in the Appendix.

## Introduction

Speech perception is influenced by both acoustic and visual information resulting from the speaker's articulatory movements. Although humans are able to understand words without visual input, auditory-visual perception has been shown to improve language recognition and comprehension (Sumby & Pollack, 1954) even when the acoustic information is clear (Reisberg, McLean, & Goldfield, 1987).

Sams et al. (1991) used magnetoencephalography (MEG), a functional neuroimaging technique that uses magnetic fields, to identify the neuroanatomical areas where auditory-visual integration occurs. Sams et al. found that visual information from articulatory movements may be used by the auditory cortex and influence auditory perception. Kislyuk, Mööttöönen, and Sams (2008) further suggested that multisensory interactions occur during early auditory processing. These results provide additional evidence that conflicting signals from auditory and visual systems merge into a unified neural representation in the early stages of sensory processing.

Saint-Amour, Sanctis, Molholm, Ritter, and Foxe (2007) likewise suggest that visual stimuli influence auditory speech perception in the auditory cortex. The use of both auditory-visual and visual only stimuli enabled the subtraction of the oddball, or deviant, visual responses from the common, or standard, stimuli responses in the auditory-visual data. No mismatch negativity (MMN), a difference waveform that is elicited by an unexpected stimulus, was observed for the visual alone condition; however, the McGurk effect showed a robust MMN response in the latency range from 175-400 ms. These results support the concept that visually driven multisensory illusory phonetic percepts are associated with an auditory MMN cortical response involving the left hemisphere temporal cortex. Hence, visual stimuli influence auditory speech perception in the auditory cortex.

Evidence has suggested that auditory-visual speech may be integrated during the early stages of speech processing (Kislyuk et al., 2008; Möttönen, Krause, Tiippana, & Sams, 2002; Pilling, 2009). Möttönen et al. (2002) investigated whether change detection mechanisms in the auditory cortex can distinguish between phonetically different unimodal visual speech stimuli, or if acoustic speech integration helps detect visual changes in the auditory cortex. By using MEG, their results showed that changes in visual speech stimuli were detected in auditory cortices, bilaterally, in the absence of acoustic stimuli. In the visual only experiment, visual changes were processed at a longer latency implying that the integration of the auditory-visual information increased the processing rate in the auditory cortex. Gentilucci and Cattaneo's (2005) work supports the hypothesis of cross-modal integration between auditory and visual inputs, as opposed to the superimposition of an automatic imitation of acoustic motor patterns on the visually detected motor patterns.

**McGurk Effect**

The McGurk effect, first described by McGurk and MacDonald in 1976, showed that a disparity between an auditory phoneme and a visual phoneme produced an auditory illusion that was either a combination of, or a bias towards, the auditory and visual phonemes. When auditory and visual stimuli conflict with each other, a new percept is observed. This new percept has been shown to be present in the phoneme contrast of /ba/ and /ga/. When the auditory stimulus is /ba/and visual stimulus is /ga/, the observer tends to fuse the two syllables. Instead of either an auditory or visual dominance, neither stimulus is favored over the other. When fusion occurs, /da/, a sound in between /ba/ and /ga/, is perceived. Inversely, when the auditory stimulus is

/ga/and visual stimulus is /ba/, the observer tends to combine the two phonemes. The perceived sound is either /bga/ or /gba/. This indicates that auditory perception can be modified by a visual stimulus. This percept is observed across a wide span of ages, from young children to adults. The results from the follow-up study by MacDonald and McGurk (1978) confirmed the predictive validity of the manner-place hypothesis, where the manner in which something is spoken is detected by audition and the place of articulation is detected by vision, with regard to the illusions that are elicited by labial-voice/nonlabial lip productions.

The McGurk effect has been a common subject of continued investigation. Green, Kuhl, Meltzo, and Stevens (1991) showed that the McGurk effect is not dependent on the gender of the speaker. The effect is equally strong when the face-voice stimuli are gender-compatible or gender-incompatible. Although the listener is aware of the details of the speaker, this information is neutralized for the task of phonetic categorization.

Fixation location can influence the perception of the McGurk effect. Paré, Richler, Hove, and Munhall (2003) suggest that the perception of this effect, and thus the integration of auditory-visual information, was not significantly enhanced by fixating on the mouth. A second experiment in their study observed that fixating within the central region (eyes and mouth) of the speaker's face preserved the McGurk effect. The third experiment demonstrated that a McGurk effect could be produced when the participant's fixation deviated superiorly as much as 40º relative to the talker's mouth. Deviations greater than 40 º did not produce the McGurk effect.

The timing of the presentation of the auditory and visual stimuli can influence the McGurk effect. While a strict synchrony of the auditory-visual stimuli is not necessary to elicit the McGurk effect, the rates of auditory and visual stimuli have a significant influence on perception (Jones & Callan, 2003; Munhall, Gribble, Sacco, & Ward, 1996; Munhall, Ten Hove, Brammer, & Paré, 2009). There is a small, but reliable, tendency for synchronous auditory-visual stimuli to produce a greater illusion.

**Neurophysiological Measures of the McGurk Effect**

**Electroencephalography.** Electroencephalography (EEG), a measure of brain electrical activity, enables the localization of neuronal populations of specific cortical regions within the central auditory nervous system (Kasai et al., 2002; Ponton, Bernstein, & Auer, 2009). The brain's electrical activity is measured by placing electrodes at specific locations across the scalp. The electrodes record ionic current flow from large populations of neurons in the brain that are activated in response to extrinsic and intrinsic stimulation (Näätänen, 1995). EEG is used to obtain event-related potentials (ERPs) from sensory stimulation (Picton 2006).

**Event-related potentials.** ERPs, a measure of sensory function, are a subset of the EEG in which a sensory stimulus is used that causes an endogenous time locked change in the EEG. According to Teplan (2002), the most useful application of the EEG recordings is the ERP. Through the identification of active neuronal populations, ERPs may be mapped across the cortex for cerebral processing with high temporal resolution (Teplan, 2002).

**Source Localization**

Source localization refers to the ability of a neural imaging technique to identify underlying neuronal activation associated with neural processing. Source localization can aid in identifying the neuronal sites that are stimulated in the auditory and visual system during the McGurk effect. Since functional magnetic resonance imaging (fMRI) gives good spatial information but poor temporal information, and Quantitative EEG (QEEG) gives good temporal information but poor spatial information, the overlaying of an MRI onto the QEEG significantly improves the prediction of the QEEG as it relates to the distribution of the sensory response across the scalp as a function of time, or temporal resolution (Dale & Halgren, 2001).

For example, Saint-Amour et al. (2007) observed that right hemispheric contributions to the McGurk MMN were accounted for with a single source in the superior temporal gyrus (STG) while two separate sources were found in the left hemisphere (in the transverse gyrus of Heschl and in the STG) using EEG and topography. Pilling (2009) discovered that the peak of the N1/P2 wave that followed the presentation of auditory-visual speech stimuli was significantly smaller in comparison to the auditory only stimuli as well as to the auditory or visual alone responses. The analysis of the data from the EEG indicated that the effect of auditory-visual speech was nonlinear. In order for an amplitude reduction to occur, the stimuli needed to be in synchrony. This finding supports the concept that this amplitude reduction effect is linked with the operation of integrative mechanisms and that some integration of auditory and visual information takes place at an early stage of processing.

An analytical technique applied to the mathematical and statistical analysis of brainwave activities (EEG) is QEEG. QEEG provides an analytical ability to view and describe dynamic changes in brainwave activity arising from exogenous and endogenous tasks leading to an estimate of cortical activities engaged in sensory and other types of brain processing. Also, QEEG permits the comparison of brainwave activity across individuals and established databases (Bagic & Sata, 2007). QEEG primarily involves measurements of power (the extent and amount of brainwave activity), spectrum (the frequency bands of the brainwave activity), asymmetry (the distribution of the power spectrum of brainwave activity), coherence (what areas of the brain are "talking" to each other), and phase (the temporal aspect of brainwave activity). Furthermore, when isopotential contours are created from these types of analyses, a map of brainwave activity across the scalp is seen (Congedo, John, De Ridder, Prichep, & Isenhart, 2010; Mathewson et al., 2012). In the current study, QEEG is used to quantify the brain's response to the presentation of stimuli.

A well-established technique for assessing and evaluating various conditions and structures of the brain by using strong magnetic fields is MRI. The images that are produced are highly sensitive and provide a good spatial resolution of the tissues of the brain (American Society of Neuroradiology, 2013; Duyn, 2012). As a result, MRI has become the imaging technique of choice used in clinical medicine and research (Duyn, 2012).

A recent transcranial magnetic stimulation (TMS) study conducted by Beauchamp, Nath, and Pasalar (2010) provided evidence for the critical role of the superior temporal sulcus (STS) in auditory-visual integration. By sending a single-pulse TMS to the participants' left STS near the time when the McGurk stimuli were presented, a significant reduction in the perception of the McGurk effect was observed. Subjects reported only the auditory component.

Nath and Beauchamp (2012) observed that the difference between McGurk perceivers and non-perceivers was found in neural responses in their left STS. Functional localizers, MRI and fMRI, were used to identify the location of the multisensory portion of STS in each participant. The McGurk susceptibility group had greater left STS activity in response to incongruent syllables than the non-perceivers of the McGurk effect. Across all participants, there was a significant positive correlation between the participants' STS response to incongruent syllables and their likelihood of experiencing the McGurk effect. No difference was found between these two groups for congruent syllables. Literature has suggested that the bilateral STS region is a major site for auditory-visual integration (Pilling, 2009; Szycik, Stadler, Tempelmann, & Munte, 2012) and is involved in the processing of the McGurk effect

(Beauchamp et al., 2010; Matchin, Groulx, & Hickok, 2014; Nath & Beauchamp, 2012). The auditory cortex plays a key role in the McGurk effect as well; the conflicting signals from auditory and visual modalities merge to form a unified neural representation during early sensory processing (Kislyuk et al., 2008; Möttönen et al., 2002). Szycik et al. (2012) also suggested that the left STS, in particular, is a key area for individual differences in speech perception. However, Jones and Callan (2003) found that no relationship was observed between perceptual performance and activation in the STS or auditory cortex. Instead, the highest amount of electrophysiological activation during incongruent stimuli presentation was located in the right supramarginal gyrus and left inferior parietal lobule.

While many studies have been conducted about the McGurk effect, none have addressed brain imaging and dipole source localization of the McGurk effect in university students majoring in the arts. This study investigated differences or similarities in how the brain processes a mismatch between auditory and visual inputs that may exist between music performance majors and visual arts majors without a musical background. It was hypothesized that the music performance majors would have a slight auditory bias, selecting what they hear over what they see when a mismatch in auditory and visual stimuli was presented. In addition it was hypothesized that there would be small differences in the locations of the dipoles between the two groups.

## Method

### Participants

Ten individuals (five per group) between the ages of 18 and 28 years participated in this study. One group consisted of two male and three female students majoring in music performance, excluding vocal performance. The second group was comprised of one male and four female students majoring in visual arts who did not have a musical background (no musical training beyond eighth grade and currently have not played a musical instrument for five years or longer). All participants were required to be native English speakers (Bomba, Choly, & Pang, 2011; Neville et al., 1998) and have no reported history of cognitive, learning, or neurological impairments (Csépe, Osman-Sági, Molnár, & Gósy, 2001). Additional qualifications required that the participants be right handed (Csépe et al., 2001), and pass an initial hearing screening showing that their hearing is within normal limits bilaterally. Hearing screening included otoscopy, tympanometry, pure tone testing, and word recognition scores. Hearing screenings met the specifications set forth by the American Speech-Language Hearing Association (ASHA, 1990). This included clear, healthy tympanic membranes bilaterally, bilateral type A tympanograms with static acoustic admittance measures between 0.3 and 1.4 mmhos, and peak pressure between -100 and +50 daPa. Normal pure tone thresholds are defined as ≤ 15 dB HL for octave intervals between 250-8000 Hz and threshold differences between ears ≤ 5 dB HL. Speech recognition thresholds (SRT) did not exceed the limits of ≤ 15 dB HL and were within 2 dB of the pure tone average. Word recognition scores were ≥ 98% bilaterally at a presentation level of 40 dB SL relative to SRT (Roup, Wiley, Safady, & Stoppenbach, 1998).

Each participant read and signed an informed consent document approved by the Institutional Review Board at Brigham Young University (see Appendix B) before participating in the study. In addition to meeting the ethical requirements set by Brigham Young University, this study also met the ethical requirements as stated in the Declaration of Helsinki (World Medical Association, 2008). The participants received compensation ($10.00 USD) for their participation in the study. Two participants from the study were chosen at random to be tested a second time in order to compare data and verify test-retest reliability.

## Instrumentation

**Stimulus preparation.** The phonemes /ba/, /da/, and /ga /were selected by recording those phonemes from a series of four college age, female, native English speakers. Recordings were completed in a sound isolated booth using a Sony MXCAM digital camcorder with a video resolution of 29.97 fps, 24 Mbps (1929 x 1080), a flat panel and a LED light array. A tripod and a remote control were utilized to minimize camera movement. The auditory stimuli were recorded using a linear 16-bit, 48 KHz sample rate and an external Sony ECM-XM1 microphone placed approximately 30 cm from the speaker's mouth using a windscreen. Three graduate students in speech-language pathology judged each speaker on a scale of one to five. The speaker with the highest score for both visual and auditory recordings was selected. The recordings for that speaker were re-evaluated for the best phoneme sample by rating each phoneme for auditory and visual clarity, again using a scale of one to five for each modality. The final phoneme selection was edited using Adobe Premiere CS5 and Adobe Audition CS5.

The phonemes were balanced for equal loudness using an adapted loudness paired-comparison paradigm (Yost, 2007). The visual aspect ratio and cropping was standardized across all visual recordings. The final visual recordings were 29.97 fps, 24 Mbps (720 x 480). The final auditory recordings were 16 bit, 48 KHz.

**Instrumentation for initial hearing screening.** Instrumentation that was used for the hearing screening included a Welch Allyn otoscope for otoscopy, a handheld Grason-Stadler model GSI-33 impedance meter for tympanometry, and a Grason-Stadler model GSI-1761 audiometer with headphones for the auditory testing. Also, during data acquisition, the test stimuli were presented to the participant via the Grason-Stadler audiometer.

Actual hearing tests were conducted in a double-walled, sound treated test booth. Noise levels were within the limits as specified by the American National Standards Institute (ANSI) S3.1-1999 R2008 Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms for ears uncovered (ANSI, 2008).

**Instrumentation for visual screening.** The participants had normal visual acuity (normal or corrected for normal vision) as measured on a Snellen® visual acuity chart. Visual acuity was not worse than 20/40 corrected in either eye. Participants were required to identify correctly all five letters on the fifth line of the Snellen® visual acuity chart while standing 20 feet away from the chart located at eye level. Each eye was tested independently.

**Stimuli**

Auditory-visual stimuli were used in this study. Three naturally spoken syllables (/ba/, /da/, and /ga/) were presented in nine paired conditions (e.g. auditory-visual) as shown in Table 1.

Table 1

*The Stimuli Conditions Presented to Each Participant*

| Type | Auditory | Visual |
|---|---|---|
| 1 | /ba/ | /ba/ |
| 2 | /ba/ | /da/ |
| 3 | /ba/ | /ga/ |
| 4 | /da/ | /da/ |
| 5 | /da/ | /ba/ |
| 6 | /da/ | /ga/ |
| 7 | /ga/ | /ga/ |
| 8 | /ga/ | /ba/ |
| 9 | /ga/ | /da/ |

**Procedure**

**Stimulus presentation and behavioral data acquisition.** The auditory and visual stimuli were placed on a personal computer (PC) and interfaced with the NeuroStim and NeuroScan software. The NeuroStim software was used to trigger the auditory and visual stimuli and to mark the stimulus type on the streaming EEG using the NeuroScan software. Twenty-three stimulus sequence files were created. Each sequence contained a randomized list of 504 phonemes, or 56 presentations of each condition. Two sequence files were randomly assigned to each participant as shown in Table 2. A total of 1008 phonemes were presented to each participant.

Table 2

*Randomized Assignment of Two Sequence Files Per Participant*

| | Sequence Files | |
|---|---|---|
| Participant | First Series | Second Series |
| 1 | 9 | 17 |
| 2 | 15 | 12 |
| 3 | 4 | 14 |
| 4 | 10 | 18 |
| 5 | 22 | 2 |
| 6 | 1 | 16 |
| 7 | 19 | 3 |
| 8 | 21 | 23 |
| 9 | 20 | 11 |
| 10 | 6 | 7 |

A Dell computer monitor was placed approximately 90 cm within the visual field of the participant. The visual presentation consisted of approximately 75% of available screen occupancy. The auditory stimulus was presented binaurally via Etymotics 3A insert earphones at 65 dB HL routed from the PC through a GSI-61 audiometer. Each stimulus included a 1200 ms epoch with an 800 ms interstimulus interval. Each of the nine phonemes was presented randomly. The participants were asked to push one of four designated response buttons: button 1 was /ba/; button 2 /da/; button 3 /ga/; and button 4 “other”. Prior to the presentation of the stimuli, participants were read the following instructions:

> You will hear a series of phonemes. You are to select one of the four buttons based upon your perception of the presented sounds. Both listen and watch; keep your eyes open. Wait for the lips to start to close before responding. For each sound you hear, firmly press the corresponding button, /ba/, /da/, or /ga/. You may also select “other.” In order for the next video to play, a button must be pressed. Press button 4, or “other”, if needed. If two videos play, respond to the second video. Please keep your gaze focused on the screen and keep as still as possible. If you wish to discontinue or pause the test at any time, you may say, “I want to stop now.” Are there any questions? We will start the test.

The experimental duration lasted 45 minutes, including a five-minute break in between the two blocks of stimuli.

**Electroencephalography data collection.** Participants sat quietly in an audiometric test room during the acquisition of the data. Participants were fitted with an electrode cap (Electro-Cap International, 2003) having 64 silver-silver chloride electrodes resting against the scalp and distributed according to the 10-20 International System (Jurcak, Tsuzuki, & Dan, 2007). In addition to the scalp electrodes, six electrodes were placed on the right and left mastoid process (linked-mastoid references), the outer canthus of the right and left eyes, and one above and below the supraorbital foramen of the left eye. These additional six electrodes were placed to monitor activity and movement of the eye and facial muscles. Electrode impedances of the cap did not exceed 3000 ohms.

Compumedics software (2008) was used for EEG data collection and initial analysis (NeuroScan 4.5). NeuroScan Stim 2 software was used for stimulus presentation. In addition, CURRY 7 (Compumedics Neuroscan, 2008) software was used for cortical localization of the electrophysiological responses, post-hoc. Participants' responses and EEG were recorded and stored on a secure digital computer.

**Data Analysis**

**Behavioral data.** The study had two independent variables and two dependent variables. The independent variables consist of the three phonemes (/ba/, /da/, and /ga/) and the participants' major (music performance or visual arts). The dependent variables consist of the participants' perception (/ba/, /da/, /ga/, or other) and latency epochs, measured in milliseconds. An ANOVA for perception was performed to determine whether the independent variables had an influence on perception. In addition, post-hoc *LSD*-tests were completed on the comparison stimuli pairs to determine whether differences from the ANOVA were significant.

**Event-related potential data.** Recordings and latency epochs were individually examined. Epochs were created from the raw EEG data. Prior to averaging the epochs, the CURRY 7 software was used to remove artifacts such as eye and jaw movement (Compumedics Neuroscan, 2008). Averages of the ERP data were calculated for each block of stimuli for each participant. Further averaging of individual ERP files for each stimulus were completed for a total of nine grand averages of the ERPs.

Dipoles, cortical source sites of electrical activity, were identified using CURRY 7 software for all individual averaged ERP files and for the grand averaged ERP files (Compumedics Neuroscan, 2008; Näätänen, 2008). Locations of each dipole were compared between groups and the grand average for all deviant and standard responses within each block of stimuli. The brain activity as measured by the electrodes from the grand averaged ERP file was used to determine the dipole locations at three latency epochs.

## Results

### Latencies

Latency epochs were broken into three groupings—early (n1), middle (n2), and late (n3). For the music group, the early epoch included latencies between 70 and 242 ms, the middle epoch between 400 and 540 ms, and the late epoch between 580 and 800 ms. For the visual arts group, the early epoch included latencies between 206 and 333 ms, the middle epoch between 445 and 586 ms, and the late epoch between 620 and 767 ms. Descriptive statistics were computed for each of these latency epochs for both the music performance and visual arts majors (Tables 3 and 4).

An ANOVA was performed for subject type measured against the nine stimulus conditions across all three latency epochs (Tables 5 and 6). Significant differences were observed during the late latency for the music performance group; $F(8, 36) = 5.816$, $p < .001$. Significant differences were also observed during the late latency for the visual arts group; $F(8, 36) = 5.113$, $p < .001$.

Post-hoc *LSD* tests were completed on the three latency epochs and the nine conditions. For the music performance group, 16 conditions showed significant differences (Table 7). None of the early latency epochs were statistically significant. With the exception of one middle latency epoch, only the late latency epochs were statistically significant. For the visual arts group, 15 conditions showed significant differences (Table 8).

Table 3

*Descriptive Statistics for Three Latency Epochs, in ms, for Each of the Nine Stimulus Presentations for Music Performance Majors*

| | Condition | *M* | *SD* | Minimum | Maximum |
|---|---|---|---|---|---|
| n1 | /ba-ba/ | 167.00 | 72.64 | 72 | 242 |
| | /ba-da/ | 170.00 | 33.49 | 116 | 205 |
| | /ba-ga/ | 126.80 | 47.76 | 90 | 209 |
| | /da-da/ | 171.40 | 48.13 | 91 | 215 |
| | /da-ba/ | 137.40 | 33.99 | 84 | 175 |
| | /da-ga/ | 163.40 | 40.67 | 95 | 198 |
| | /ga-ga/ | 136.20 | 47.83 | 74 | 207 |
| | /ga-ba/ | 155.40 | 61.02 | 75 | 211 |
| | /ga-da/ | 142.60 | 62.37 | 74 | 226 |
| | Total | 152.24 | 49.10 | 72 | 242 |

| | Condition | *M* | *SD* | Minimum | Maximum |
|---|---|---|---|---|---|
| n2 | /ba-ba/ | 470.20 | 52.39 | 402 | 539 |
| | /ba-da/ | 479.00 | 47.13 | 418 | 527 |
| | /ba-ga/ | 483.40 | 29.87 | 441 | 520 |
| | /da-da/ | 502.80 | 23.76 | 475 | 530 |
| | /da-ba/ | 483.00 | 23.51 | 445 | 506 |
| | /da-ga/ | 479.40 | 22.06 | 441 | 495 |
| | /ga-ga/ | 490.20 | 38.21 | 445 | 533 |
| | /ga-ba/ | 456.00 | 36.20 | 407 | 502 |
| | /ga-da/ | 484.40 | 29.37 | 456 | 526 |
| | Total | 480.93 | 34.05 | 402 | 539 |

| | Condition | *M* | *SD* | Minimum | Maximum |
|---|---|---|---|---|---|
| n3 | /ba-ba/ | 737.00 | 46.22 | 679 | 799 |
| | /ba-da/ | 715.20 | 20.14 | 690 | 740 |
| | /ba-ga/ | 709.20 | 29.72 | 670 | 744 |
| | /da-da/ | 716.80 | 23.46 | 691 | 753 |
| | /da-ba/ | 733.00 | 29.99 | 707 | 780 |
| | /da-ga/ | 690.00 | 24.18 | 668 | 726 |
| | /ga-ga/ | 710.80 | 20.14 | 686 | 729 |
| | /ga-ba/ | 647.40 | 50.63 | 587 | 698 |
| | /ga-da/ | 638.20 | 33.25 | 599 | 678 |
| | Total | 699.73 | 44.58 | 587 | 799 |

*Note*. n1 = early latency; n2 = middle latency; n3 = late latency. The auditory phoneme is displayed first followed by the visual phoneme.

Table 4

*Descriptive Statistics for Three Latency Epochs, in ms, for Each of the Nine Stimulus Presentations for Visual Arts Majors*

| | Condition | *M* | *SD* | Minimum | Maximum |
|---|---|---|---|---|---|
| n1 | /ba-ba/ | 267.60 | 32.68 | 238 | 309 |
| | /ba-da/ | 262.40 | 11.97 | 246 | 275 |
| | /ba-ga/ | 247.20 | 33.64 | 207 | 286 |
| | /da-da/ | 271.00 | 20.14 | 251 | 292 |
| | /da-ba/ | 287.40 | 28.40 | 263 | 333 |
| | /da-ga/ | 263.40 | 33.04 | 206 | 288 |
| | /ga-ga/ | 267.80 | 20.36 | 246 | 295 |
| | /ga-ba/ | 270.40 | 38.06 | 223 | 323 |
| | /ga-da/ | 270.60 | 25.28 | 235 | 300 |
| | Total | 267.53 | 27.39 | 206 | 333 |

| | Condition | *M* | *SD* | Minimum | Maximum |
|---|---|---|---|---|---|
| n2 | /ba-ba/ | 512.80 | 36.16 | 458 | 552 |
| | /ba-da/ | 521.20 | 39.66 | 462 | 554 |
| | /ba-ga/ | 489.80 | 41.95 | 445 | 543 |
| | /da-da/ | 502.00 | 15.10 | 484 | 526 |
| | /da-ba/ | 520.00 | 37.30 | 470 | 558 |
| | /da-ga/ | 506.40 | 35.11 | 474 | 551 |
| | /ga-ga/ | 501.80 | 16.24 | 487 | 520 |
| | /ga-ba/ | 510.60 | 23.06 | 481 | 544 |
| | /ga-da/ | 559.00 | 25.10 | 524 | 586 |
| | Total | 513.73 | 34.04 | 445 | 586 |

| | Condition | *M* | *SD* | Minimum | Maximum |
|---|---|---|---|---|---|
| n3 | /ba-ba/ | 709.60 | 24.53 | 693 | 753 |
| | /ba-da/ | 721.80 | 39.78 | 668 | 767 |
| | /ba-ga/ | 692.20 | 15.09 | 678 | 715 |
| | /da-da/ | 630.20 | 8.96 | 620 | 643 |
| | /da-ba/ | 702.00 | 25.03 | 667 | 730 |
| | /da-ga/ | 719.60 | 33.07 | 674 | 759 |
| | /ga-ga/ | 709.80 | 22.86 | 676 | 739 |
| | /ga-ba/ | 711.80 | 29.10 | 678 | 750 |
| | /ga-da/ | 706.80 | 34.71 | 678 | 759 |
| | Total | 700.42 | 36.30 | 620 | 767 |

*Note*. n1 = early latency; n2 = middle latency; n3 = late latency. The auditory phoneme is displayed first followed by the visual phoneme.

Table 5

*Summary of ANOVA Results for the Music Performance Group Across the Three Latency Epochs*

| Condition | *SS* | *df* | *MS* | *F* |
|---|---|---|---|---|
| n1 | | | | |
| Between | 11,262.71 | 8 | 1,407.84 | 0.535 |
| Within | 94,819.60 | 36 | 2,633.88 | |
| Total | 106,082.31 | 44 | | |
| n2 | | | | |
| Between | 6,646.80 | 8 | 830.85 | 0.674 |
| Within | 44,376.00 | 36 | 1,232.67 | |
| Total | 51,022.80 | 44 | | |
| n3 | | | | |
| Between | 49,289.60 | 8 | 6,161.20 | 5.816* |
| Within | 38,135.20 | 36 | 1,059.31 | |
| Total | 87,424.80 | 44 | | |

*Note*. n1 = early latency; n2 = middle latency; n3 = late latency.
* $p < .001$

Table 6

*Summary of ANOVA Results for the Visual Arts Group Across the Three Latency Epochs*

| Condition | *SS* | *df* | *MS* | *F* |
|---|---|---|---|---|
| n1 | | | | |
| Between | 4,406.40 | 8 | 550.80 | 0.693 |
| Within | 28,596.80 | 36 | 794.36 | |
| Total | 33,003.20 | 44 | | |
| n2 | | | | |
| Between | 15,307.20 | 8 | 1,913.40 | 1.931 |
| Within | 35,669.60 | 36 | 990.82 | |
| Total | 50,976.80 | 44 | | |
| n3 | | | | |
| Between | 30,841.78 | 8 | 3,855.22 | 5.113* |
| Within | 27,145.20 | 36 | 754.03 | |
| Total | 57,986.98 | 44 | | |

*Note*. n1 = early latency; n2 = middle latency; n3 = late latency.
* $p < .001$

Table 7

*Comparisons of the Conditions with Statistically Significant Latency Epochs, p < .05, Within the Music Performance Group*

| Condition | /da-da/ | /da-ba/ | /da-ga/ | /ga-ba/ | /ga-da/ |
|---|---|---|---|---|---|
| /ba-ba/ | | | n3 | | n3 |
| /ba-da/ | | | | n3 | n3 |
| /ba-ga/ | | | | n3 | n3 |
| /da-da/ | | | | n2, n3 | n3 |
| /da-ba/ | | | n3 | n3 | n3 |
| /da-ga/ | | | | n3 | n3 |
| /ga-ga/ | | | | n3 | n3 |
| /ga-ba/ | | | | | |

*Note*. n1 = early latency; n2 = middle latency; n3 = late latency. The auditory phoneme is displayed first followed by the visual phoneme.

Table 8

*Comparisons of the Conditions with Statistically Significant Latency Epochs, p < .05, Within the Visual Arts Group*

| Condition | /da-da/ | /da-ba/ | /da-ga/ | /ga-ba/ | /ga-da/ |
|---|---|---|---|---|---|
| /ba-ba/ | n3 | | | | n2 |
| /ba-da/ | n3 | | | | |
| /ba-ga/ | n3 | n1 | | | n2 |
| /da-da/ | | n3 | n3 | n3 | n2, n3 |
| /da-ba/ | | | | | |
| /da-ga/ | | | | | n2 |
| /ga-ga/ | n3 | | | | n2 |
| /ga-ba/ | | | | | n2 |

*Note*. n1 = early latency; n2 = middle latency; n3 = late latency. The auditory phoneme is displayed first followed by the visual phoneme.

**Repeated Measures**

To determine test-retest reliability, ERPs were acquired a second time from two participants following the initial testing. Individual and grand averaged ERP waveform files were created for each test-retest participant. A *t*-test was conducted on the grand averaged ERP waveform files. No significant differences, $t(2) = 1.76$, $p > 0.05$, were found in either participant for repeated measures, which indicates that an acceptable level of test-retest reliability was established.

**Quantitative EEG Figures**

Figure 1 shows the distribution of the electrical brain activity across the scalp for the nine stimulus conditions and the three time epochs for the two groups of participants. Figure 2 is an axial view of the brain map depicting the same conditions across the same time epochs. In both figures, a green dot marks the approximate location of the dipole.

The brain maps for the n1 epoch across all nine stimulus conditions for the music performance group show activity primarily in the mid-temporal to mid-central areas. Source localization (Figure 2) is generally seen in the areas of the precentral gyrus of the left parietal lobe. In addition, areas in the left occipital regions are generally activated for each of the nine conditions. In contrast, the visual arts group appears to show primary activation in the left temporal areas (Figure 1), and dipole source localization (Figure 2) is seen primarily in the left lateral areas of the precentral gyrus.

The n2 epoch for the brain map in the music performance group shows general electrical activation in the right frontal areas for the /ba-ba/, /ba-da/, /ba-ga/, /da-ga/, /ga-ba/, and /ga-da/ conditions (six out of nine). Conditions /da-da/, /da-ba/, and /ga-ga/ have greater activity in the mid-posterior areas of the frontal lobe and the mid-anterior region of the parietal lobe near the central part of the central sulcus (Figures 1 and 2). In comparison, in the visual arts group, the primary distribution of electrical activity as seen in the brain map is in the left temporal lobe.

The n3 epoch illustrates for both groups primarily brain activity in the frontal areas (Figure 1). Dipole source localization (Figure 2) is primarily located in the posterior cingulate.

The exception to this is in the /da-da/ and /da-ba/ conditions for the visual group, which is seen in the superior frontal gyrus (Figure 2). The n3 epoch has latency values consistent with cognitive and perhaps some syntactic or semantic processing.

Tables 9 and 10 are a summary of the distribution of the latency times with averages and sources taken from the grand average across each group for each condition.

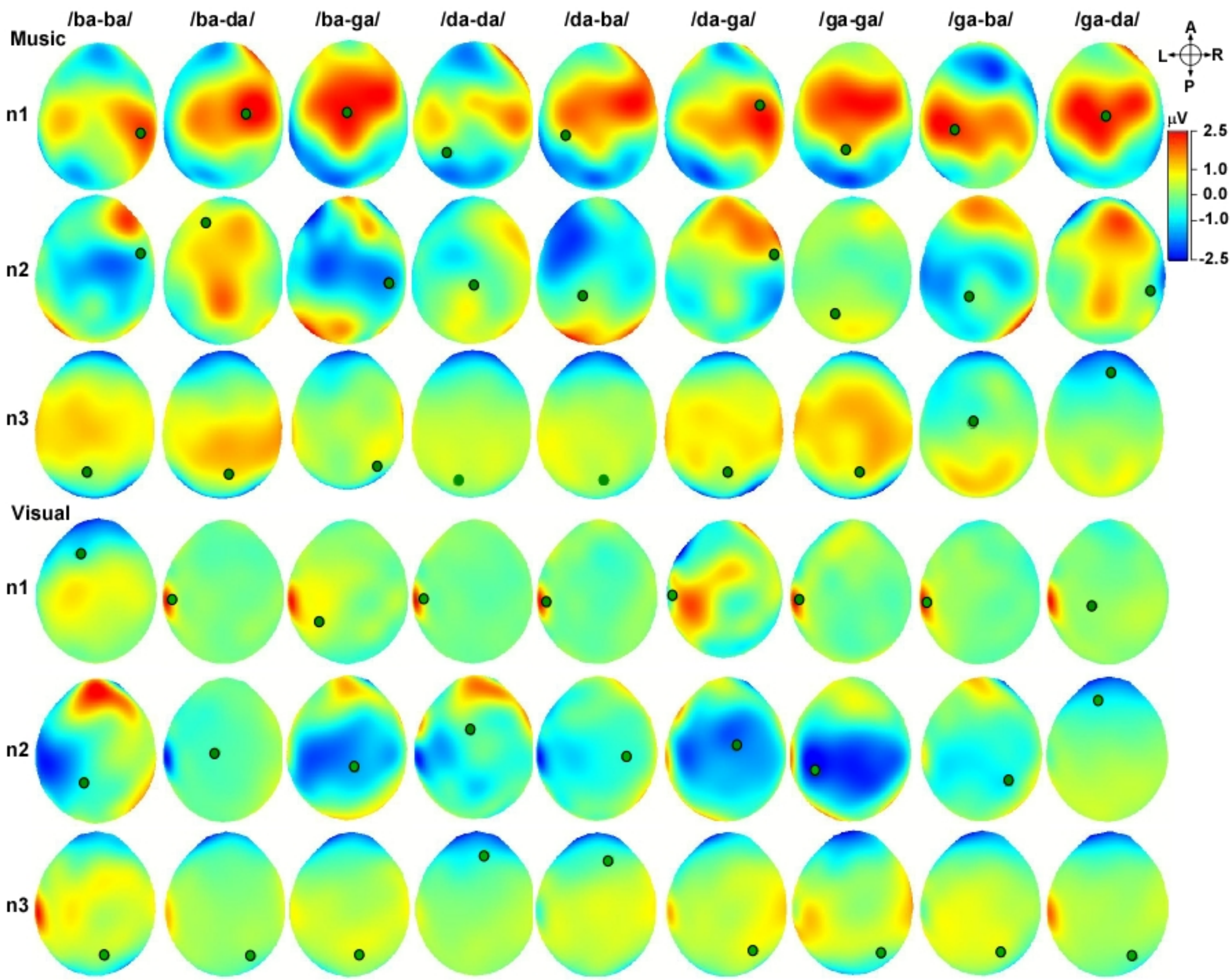


*Figure 1*. Axial view of the grand-averaged brain maps of the event related potentials for music performance and visual arts majors across all three latency epochs (early = n1; middle = n2; late = n3) for each of the nine conditions. The auditory stimuli are displayed first and then the visual stimuli. The green dot marks the approximate location of the dipole.

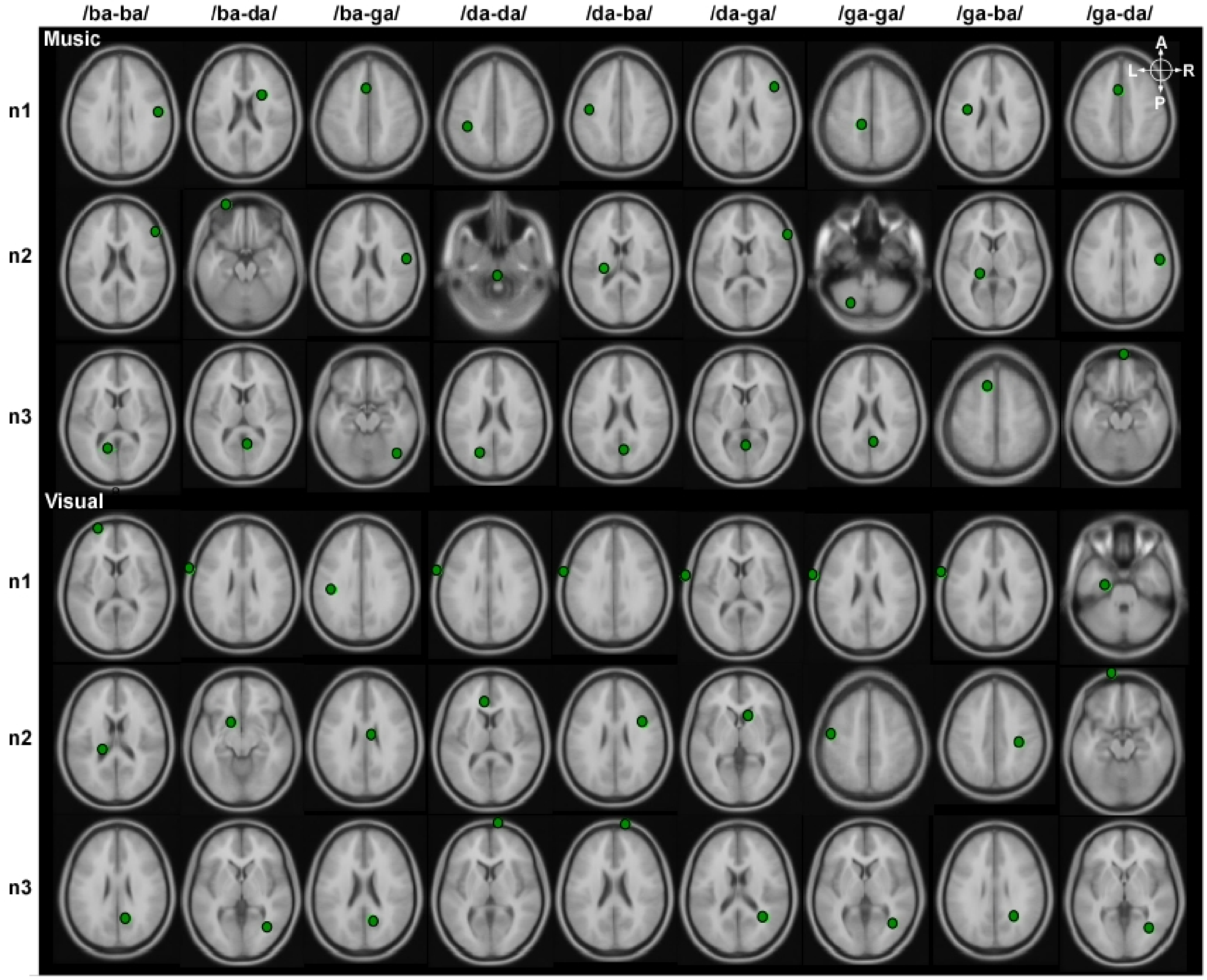


*Figure 2*. Axial spatial view of the grand-averaged brain maps of the dipole source locations for music performance and visual arts majors across all three latency epochs (early = n1; middle = n2; late = n3) for each of the nine conditions. The auditory stimuli are displayed first and then the visual stimuli. The green dot marks the approximate location of the dipole.

Table 9

*Dipole Source Locations for the Music Group for Each of the Conditions for the Three Latency Epochs from Earliest Latency to the Latest Latency Time*

| Condition | Latency (ms) | Dipole Source | Lobe |
|---|---|---|---|
| /ga-da/ | 126 | Cingulate Gyrus | Medial Frontal |
| /ba-ga/ | 129 | Medial Frontal Gyrus | Medial Frontal |
| /ga-ga/ | 136 | Paracentral Lobule | Left Posterior Frontal |
| /da-ba/ | 148 | Precentral Gyrus | Left Temporal |
| /ba-da/ | 163 | Middle Frontal Gyrus | Right Frontal |
| /da-ga/ | 180 | Middle Frontal Gyrus | Right Frontal |
| /da-da/ | 194 | Postcentral Gyrus | Left Temporal/Parietal |
| /ba-ba/ | 195 | Precentral Gyrus | Right Temporal |
| /ga-ba/ | 196 | Extra-nuclear | Left Temporal |
| /ba-da/ | 455 | Superior Frontal Gyrus | Left Frontal |
| /ga-ba/ | 455 | Thalamus- Pulvinar | Cerebral cortex/Midbrain |
| /da-ga/ | 465 | Inferior Frontal Gyrus | Right Frontal |
| /da-ba/ | 477 | Thalamus- Pulvinar | Cerebral cortex/Midbrain |
| /ba-ba/ | 478 | Middle Frontal Gyrus | Right Frontal |
| /ba-ga/ | 479 | Pre/post Central Gyrus | Right Temporal |
| /ga-da/ | 486 | Pre/post Central Gyrus | Right Temporal |
| /da-da/ | 487 | Left Anterior Thalamus | Cerebral cortex/Midbrain |
| /ga-ga/ | 488 | Inferior Semi-Lunar Lobule | Left Occipital |
| /ga-ba/ | 615 | Superior Frontal Gyrus | Medial Frontal |
| /ga-da/ | 645 | Superior Frontal Gyrus | Medial Frontal |
| /da-ba/ | 712 | Posterior Cingulate | Right Occipital |
| /ba-ga/ | 714 | Fusiform Gyrus | Right Occipital |
| /ga-ga/ | 715 | Posterior Cingulate | Right Occipital |
| /ba-da/ | 722 | Posterior Cingulate | Right Occipital |
| /da-ga/ | 724 | Posterior Cingulate | Medial Occipital |
| /da-da/ | 726 | Posterior Cingulate | Left Occipital |
| /ba-ba/ | 746 | Posterior Cingulate | Left Occipital |

Table 10

*Dipole Source Locations for the Visual Arts Group for Each of the Conditions for the Three Latency Epochs from Earliest Latency to the Latest Latency Time*

| Condition | Latency (ms) | Dipole Source | Lobe |
|---|---|---|---|
| /ba-ga/ | 263 | Postcentral Gyrus | Left Temporal/Parietal |
| /ga-ba/ | 263 | Precentral Gyrus | Left Temporal |
| /da-ga/ | 268 | Precentral Gyrus | Left Temporal |
| /ba-da/ | 277 | Precentral Gyrus | Left Temporal |
| /ga-ga/ | 283 | Precentral Gyrus | Left Temporal |
| /da-ba/ | 284 | Precentral Gyrus | Left Temporal |
| /ga-da/ | 286 | Parahippocampal Gyrus | Left Temporal/Occipital |
| /da-da/ | 287 | Precentral Gyrus | Left Temporal |
| /ba-ba/ | 290 | Precentral Gyrus | Left Frontal |
| /ba-ga/ | 481 | Cingulate Gyrus | Medial Parietal |
| /ga-ga/ | 494 | Precentral Gyrus | Left Temporal |
| /ga-ba/ | 497 | Postcentral Gyrus | Right Temporal |
| /da-da/ | 502 | Anterior Cingulate | Medial Frontal |
| /ba-ba/ | 504 | Thalamus- Pulvinar | Cerebral cortex/ Midbrain |
| /da-ga/ | 511 | Caudate | Right Parietal |
| /da-ba/ | 525 | Precentral Gyrus | Right Temporal |
| /ba-da/ | 538 | Caudate | Left Parietal |
| /ga-da/ | 651 | Superior Central Gyrus | Left Frontal |
| /da-da/ | 661 | Superior Frontal Gyrus | Right Frontal |
| /ba-ga/ | 697 | Posterior Cingulate | Right Occipital |
| /ba-ba/ | 701 | Posterior Cingulate | Right Occipital |
| /ga-ga/ | 713 | Middle Occipital Gyrus | Right Occipital |
| /ga-da/ | 714 | Middle Occipital Gyrus | Right Occipital |
| /ga-ba/ | 726 | Posterior Cingulate | Right Temporal |
| /ba-da/ | 729 | Middle Occipital Gyrus | Right Occipital |
| /da-ba/ | 732 | Superior Frontal Gyrus | Right Frontal |
| /da-ga/ | 739 | Superior Temporal Gyrus/ | Right Occipital |

## Discussion

The purpose of the current study was to determine whether distinct differences exist between those who are music performance majors, who spend many hours immersed in music and stimulation of their sense of hearing, and visual arts majors, who hone their sense of sight, when they are presented with mismatched auditory and visual stimuli. In particular, this study looked at the processing of the McGurk effect between the two groups. Processing similarities and differences are seen in electrophysiological responses (i.e., ERP latencies and dipole locations) within and between these two groups.

**Summary and Evaluation of Results**

Measurements of latency across three time epochs provided important information about when the processing of the various conditions took place. In the music performance group, the n1 latency epochs occur about 100 ms before the n1 latency epochs in the visual arts group. However, the n3 latency epochs are comparable in time (Tables 3 and 4). This finding shows that sensory integration takes about the same amount of time overall between the two groups. While the visual group takes longer to process this information in the early epochs, where perceptual events initially are being separated, the processing does not continue to lag behind. Once sorted out, the sensory integration takes about the same amount of time to merge in the brain. Other differences exist between the two groups.

For those in the music performance group, the McGurk condition, /ba-ga/, has an earlier mean latency time than the matched condition /da-da/ (Table 3). More specifically, it has the smallest mean time (126.80 ms compared to 171.40 ms) relative to the visual arts group. The largest *SD*s are observed in the n1 epoch (47.76 ms for /ba-ga/ and 48.13 ms for /da-da/). For the n1 and n2 epochs, /da-da/ has the latest mean latency time out of all of the nine conditions (171.40 ms and 502.80 ms, respectively). In the n3 epochs, the standard deviations remain the same as the n2 epochs for these two conditions (around 29.75 ms for /ba-ga/ and around 23.60 ms for /da-da/). Although the mean time difference between these two conditions decreases with each latency measure, the mean time for /da-da/ consistently remains longer. This signifies that

/da-da/ requires a longer processing time than /ba-ga/ and that the variance decreases from the n1 epoch to the n2 epoch.

For those in the visual arts group, the /da-da/ condition consistently has less variance than the McGurk condition, /ba-ga/, across the three latency epochs (Table 4). The variance for the former also decreases over time (*SD* at the n1 epoch of 20.14 to 15.10 to 8.96). In the McGurk condition, the variance increases at the n2 epoch (*SD* = 41.95), which is the largest variance observed among all of the conditions, although it has the smallest mean latency time (489.80 ms). This signifies that there is greater variation amongst the participants for the middle latency epoch. During the n3 epoch the variance significantly is reduced (*SD* = 15.09) which shows more consistency within the visual arts group. At the n1 and n2 epochs the McGurk condition has smaller mean latency times (247.20 ms and 489.80 ms) than the matched /da-da/ condition (271.00 ms and 502.00 ms). At the n3 epoch, this is reversed (/da-da/ at 630.20 ms and /ba-ga/ at 692.20 ms). In addition, the visual arts group has the shortest mean latency time at the n3 epochs for the /da-da/ condition (630.20 ms) compared to the music performance group (716.80 ms), meaning that, collectively, the visual arts group is able to process this condition quicker than the music performance group. This shows that, initially, the visual arts group begins to process the McGurk effect earlier, but cognition, which occurs around the n3 epoch, requires more processing time when compared to the matched /da-da/ condition.

In both groups no statistically significant differences relative to latency are seen across the nine conditions in the n1 and n2 epochs. In contrast, the n3 epoch for both groups shows significant differences across epochs (Tables 5 and 6). This would suggest that the point at which processing becomes biased between the auditory or visual stimuli occurs during advanced cognition; decisions are made at the later time (the n3 epoch).

For the music group, the auditory /ga/ produces the most confusion because there are differences in processing when it is an auditory /ga/ and visual /ga/. A post-hoc test for the music performance group shows that the greatest confusions occur when participants are differentiating between auditory /ga/ and other visual conditions (Table 7). Only one n2 epoch condition has a statistically significant difference (/da-da/ with /ga-ba/) while the remainder occur in the n3 epoch. In addition, the McGurk effect condition /ba-ga/ significantly differs with /ga-ba/ and /ga-da/ at the n3 epoch.

For the visual group, similar trends are observed but with the phoneme /da/ (Table 8). Earlier processing differences are seen within visual arts group as there are statistically significant differences within the earlier latency epochs. The McGurk effect condition /ba-ga/ significantly differs with /da-ba/ at the n1 epoch, /ga-da/ at the n2 epoch, and /da-da/ at the n3 epoch (Table 8). Perceptual differences start from the n1 epoch. While the conditions vary, a common element is the presentation of the auditory /da/. These results regarding the /ba-ga/ condition show that there are differences in how the music performance majors and visual arts majors process the McGurk effect.

At the n1 epoch, source localization (Figure 2) for the music performance group is seen in the general areas of the precentral gyrus of the left parietal lobe. This area is located near the primary motor cortex, Broca's area. Much of this area is associated with motor speech production as well as the analysis of auditory and visual articulation of speech (Bookheimer, 2002; Matchin et al., 2014). Likewise, we see areas in the left occipital regions generally activated for each of the nine conditions. This would be consistent with activation of the visual cortex (Sams et al., 1991). Of interest, the visual group appears to show primary activation in the left temporal areas (Figure 1) for the n1 epoch. Likewise, dipole source localization (Figure 2) is primarily seen in the left lateral areas of the precentral gyrus. The lack of activation of the occipital areas and consistent lateralization of the dipoles to the left temporal region for the visual group strongly suggest that in decoding the perceptual aspects of the stimuli they are heavily dependent upon the auditory areas of the left temporal lobe. It should be noted that these areas are located near Wernicke's area thus suggesting, in addition to auditory perceptual decoding, attempts to attach the decoded language to semantic processing (Bookheimer, 2002). At the n2 epoch for the visual arts group, it is noted again that the primary distribution of electrical activity is in the left temporal lobe as seen in the brain map. This suggests that the visual group depends more on auditory-linguistic information than the music group.

In the music performance group, the source locations of the dipoles are in the frontal and temporal lobe during the n1 epoch. The majority of the dipoles shift to the occipital lobe during the later latency epochs. In the visual arts group, the source locations of the dipoles are in the left temporal lobe during the early latency epoch. Saint-Amour et al. (2007) also found left lateralization at the early latency epoch. At the n3 epoch, the majority of the dipoles shift from the left temporal lobe to the right occipital lobe. Qualitatively, Figure 1 shows more consistent brain mapping within each condition for the visual arts group than for the music performance group. That is, it appears that the music performance group has a more complex decoding system than the visual arts group. One possibility is that the music performance group depends more on spectral content of the phoneme pairs than the visual arts group. It also may suggest that the visual arts group is able to better integrate the visual and auditory information to resolve the conflict.

In both groups, two conditions show dipole locations in the frontal lobe during the n3 epoch. However, the conditions vary by group. For those in the music performance group, visually mismatched conditions /ga-ba/ and /ga-da/ have dipoles in the superior frontal gyrus. In the visual arts group, the two conditions with activity in the superior frontal gyrus are /da-da/ and /da-ba/. The superior frontal gyrus is involved in various cognitive and motor control tasks as well as working memory and attention (Li et al., 2013). More differences are seen in the processing of the /ga/ phoneme in the music performance group while more differences are seen in the processing of the /da/ phoneme in the visual arts group.

Many researchers have observed that the STS plays an integral role in the integration of auditory and visual information (Beauchamp et al., 2010; Campbell, 2008; Matchin et al., 2014; Nath & Beauchamp, 2012; Pilling, 2009; Szycik et al., 2012). The results of the current investigation show more specific activity within the pSTS (Tables 9 and 10). Other studies have shown that cross modal integration occurs between auditory and visual inputs (Gentilucci & Cattaneo, 2005) from an automatic, pre-cognitive comparison during early auditory processing (Colin et al., 2002; Kislyuk et al., 2008). This study provides additional evidence in support of this observation. Neither the matched nor the mismatched pairs, as a whole, were processed quicker than the other (Tables 3 and 4). The results were mixed, especially in the visual arts group. Both modalities are integrated early on. This result supports the finding of Campbell (2008), who reported that a combined auditory and visual processing is most effective in processing natural speech.

**Limitations and Recommendations for Future Research**

A limitation of this study is in the inability to have two distinct groups without confounding variables. It is near improbable for there to be a complete divide between the two groups, where one group rarely uses a sense, either hearing or sight, when both senses work well. For those individuals without hearing impairment, sound is heard everywhere, including spoken words and music. In addition, for those individuals without visual impairment, sight is inevitable. Eyes are used to see, including the faces of those who speak. The inclusion requirements to participate in this study were used to reduce these confounding variables.

Future investigation may want to expand upon the current study by examining the behavioral reaction times for the matched versus the mismatched stimuli. Further analysis may provide additional insight into the brain's processing of the mismatched stimuli (Nahorna, Berthommier, & Schwartz, 2012). It may also be informative to analyze the hit/miss results for each participant to see if a preference exists for either the auditory or visual signal. The nature of the MMN response may be examined as well (Colin et al., 2002; Kislyuk et al., 2008; Saint-Amour et al., 2007).

To investigate whether the environment can influence the processing of auditory and visual signals or if brains are prewired certain ways that then influences the field of study that is pursued, a longitudinal study can compare the changes in the brain activity over time, from elementary days, to high school, and then higher education. From such data brain plasticity can be observed. In addition, researchers can discover if those who are musically inclined show stronger auditory preferences from a young age or if this preference develops with more intensive training over the years.

**References**

American National Standards Institute. (2008). *American national standard: Maximum permissible ambient noise levels for audiometric test rooms*. ANSI S3.1-1999 R2008. New York, NY: ANSI.

American Society of Neuroradiology (2013). ACR-ASNR Practice Guideline for the Performance and Interpretation of Magnetic Resonance Imaging (MRI) of the Brain. *Resolution, 6*, 1-16. Retrieved from http://www.acr.org

American Speech-Language-Hearing Association. (1990). Guidelines for screening for hearing impairments and middle-ear disorders. *American Speech-Language-Hearing Association, 32*(2), 17-24. Retrieved from http://www.asha.org

Bagic, A., & Sata, S. (2007). Principles of electroencephalography and magnetoencephalography. In F. Hillary, & J. DeLuca (Eds.), *Functional neuroimaging in clinical populations* (pp. 71-96). London, England: Guilford Press.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30*, 2414-2417. doi: 10.1523/JNEUROSCI.4865-09.2010

Bomba, M. D., Choly, D., & Pang, E. W. (2011). Phoneme discrimination and mismatch negativity in English and Japanese speakers. *Neuroreport 22*(10), 479-483. doi: 10.1097/WNR.0b013e328347dada

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience, 25*(1), 151–188. doi: 10.1146/annurev.neuro.25.112701.142946

Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society*, *363*, 1001–1010. doi: 10.1098/rstb.2007.2155

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology, 113*, 495-506. doi:10.1016/S1388-2457(02)00024-X

Compumedics Neuroscan (2008). CURRY 7 [computer software]. North Carolina: Compumedics USA.

Congedo, M., John, R. E., De Ridder, D., Prichep, L., & Isenhart, R. (2010). On the "dependence" of "independent" group EEG sources: An EEG study on two large databases. *Brain Topography, 23*(2), 134-138. doi: 10.1007/s10548-009-0113-6

Csépe, V., Osman-Sági, J., Molnár, M., & Gósy, M. (2001). Impaired speech perception in aphasic patients: Event-related potential and neuropsychological assessment. *Neuropsychologia, 39*(11), 1194-1208. doi: 10.1016/s0028-3932(01)00052-5

Dale, A. M., & Halgren, E. (2001). Spatiotemporal mapping of brain activity by integration of multiple imaging modalities. *Current Opinion in Neurobiology*, 11, 202-228.

Duyn, J. H. (2012). The future of ultra-high field MRI and fMRI for study of the human brain. *Neuroimage, 62*(2), 1241-1248. doi: 10.1016/j.neuroimage.2011.10.065

Electro-Cap International, I. (2003). *Electro-cap [equipment]*. Eaton, OH: Electro-Cap International, Inc.

Gentilucci, M., & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Experimental Brain Research, 167*(1), 66-75. doi: 10.1007/s00221-005-0008-z

Green, K. P., Kuhl, K. P., Meltzo, N. A., & Stevens, E. R. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics, 50*, 524-536.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Neuroreport, 14*, 1129-1133.

Jurcak, V., Tsuzuki, D., & Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *Neuroimage, 34*(4), 1600-1611. doi:10.1016/j.neuroimage.2006.09.024

Kasai, K., Nakagome, K., Iwanami, A., Fukuda, M., Itoh, K., Koshida, I., & Kato, N. (2002). No effect of gender on tonal and phonetic mismatch negativity in normal adults assessed by a high-resolution EEG recording. *Cognitive Brain Research, 13*(3), 305-312. doi: 10.1016/s0926-6410(01)00125-2

Kislyuk, D. S., Mööttöönen, R., & Sams, M. (2008). Visual processing affects the neural basis of auditory discrimination. *Journal of Cognitive Neuroscience, 20*, 2175-2184. doi: 10.1162/jocn.2008.20152

Li, W., Qin, W., Liu, H., Fan, L., Wang, J., Jiang, T., & Yu, C. (2013). Subregions of the human superior frontal gyrus and their connections. *Neuroimage, 78*, 46-58. doi:10.1016/j.neuroimage.2013.04.011

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*, 253-257.

Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: Evidence from articulatory suppression, the McGurk Effect, and fMRI. *Journal of Cognitive Neuroscience*. *26*(3), 606–620. doi:10.1162/jocn_a_00515

Mathewson, K. J., Jetha, M. K., Drmic, I. E., Bryson, S. E., Goldberg, J. O., & Schmidt, L. A. (2012). Regional EEG alpha power, coherence, and behavioral symptomatology in autism spectrum disorder. *Clinical Neurophysiology, 123*(9), 1798-1809. doi: 10.1016/j.clinph.2012.02.061

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

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research, 13*(3), 417-425. doi: 10.1016/S0926-6410(02)00053-8

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58*, 351-362.

Munhall, K. G., Ten Hove, M. W., Brammer, M., & Paré, M. (2009). Audiovisual integration of speech in a bistable illusion. *Current Biology, 19*, 735-739. doi: 10.1016/j.cub.2009.03.019

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing, 16*(1), 6-18. doi: 10.1097/00003446-199502000-00002

Näätänen, R. (2008). Mismatch negativity (MMN) as an index of central auditory system plasticity. *International Journal of Audiology, 47*(Supplement 2), S16-20. doi: 10.1080/14992020802340116

Nahorna, O., Berthommier, F., & Schwartz, J. (2012). Binding and unbinding the auditory and visual streams in the McGurk effect. *Journal of the Acoustical Society of America, 132*, 1061-1077.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage, 59*(1), 781-787. doi: 10.1016/j.neuroimage.2011.07.024

Neville, H. J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A., . . . Turner, R. (1998). Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. *Proceedings of the National Academy of Sciences of the United States of America, 95*(3), 922-929. doi: 10.1073/pnas.95.3.922

Paré, M., Richler, R., Hove, M., & Munhall, K.G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception and Psychophysics, 65*, 553-567. doi: 10.3758/BF03194582

Picton, T. W. (2006). Auditory event-related potentials. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science.* Hoboken, NJ: John Wiley & Sons, Ltd. doi: 10.1002/0470018860.s00486

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *Journal of Speech, Language, and Hearing Research, 52*, 1073-1081. doi: 1092-4388/09/5204-1073

Ponton, C. W., Bernstein, L. E., and Auer, E. T. (2009). Mismatch negativity with visual-only and audiovisual speech. *Brain Topography, 21*, 207-215. doi 10.1007/s10548-009-0094-5

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but not to understand: A lipreading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). Hillsdale, NJ: Erlbaum.

Roup, C. M., Wiley, T. L., Safady, S. H., & Stoppenbach, D. T. (1998). Tympanometric screening norms for adults. *American Journal of Audiology, 7*, 55–60. doi: 10.1044/1059-0889(1998/014)

Saint-Amour, D., Sanctis, P. D., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*, 587-597. doi: 10.1016/j.neuropsychologia.2006.03.036

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S .T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*, 141–145. doi: 10.1016/0304-3940(91)90914-F

Sumby, W. H., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.

Szycik, G., Stadler, J., Tempelmann, C., & Munte, T. (2012). Examining the McGurk illusion using high-field 7 tesla functional MRI. *Frontiers in Human Neuroscience, 6,* 1-7. doi: 10.3389/fnhum.2012.00095

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review, 2*(2), 1-11.

World Medical Association (2008). *WMA declaration of Helsinki: Ethical principles for medical research involving human subjects.* World Medical Association, Inc. Retrieved from http://www.wma.net/en/30publications/10policies/b3/index.html

Yost, W. A. (2007). *Fundamentals of hearing*. New York, NY: Elsevier.

**Appendix A**

**Annotated Bibliography**

American Speech-Language-Hearing Association (1990). Guidelines for screening for hearing impairments and middle-ear disorders. *American Speech-Language-Hearing Association, 32*(2), 17-24. Retrieved from http://www.asha.org

*Objective:* The American Speech-Language Hearing Association (ASHA) publishes specific guidelines regarding screening and assessing individuals for hearing impairments and disorders. These guidelines are set forth to safeguard against unethical practice in conducting hearing screenings. In addition, these guidelines ensure that results of hearing screenings are interrupted the same nationwide. *Relevance to current work:* Each participant in the current study had a hearing screening in order to be considered for additional QEEG investigation. The guidelines set forth by ASHA were followed in the participants' initial hearing screenings. *Level of evidence:* N/A.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 30*, 2414-2417. doi: 10.1523/JNEUROSCI.4865-09.2010

*Objective:* This study was designed to show that the STS is involved in the processing of the McGurk effect by combining fMRI and TMS. *Study Sample:* Twelve participants (mean age 25 years) composed this study. *Methods:* In experiment 1 and 3, a male speaker was recorded saying "ba" and "ga" while in experiment 2 a female speaker was recorded saying "pa", "ka" and "na". The subject completed two runs, one with the TMS coil targeting the left STS and one with the TMS coil targeting a control site. Single-pulse TMS was delivered to the STS at one of 11 times with a longer latency before or after the auditory onset was presented. MRI and fMRI were also used to take into consideration individual differences. *Results:* The main location of activity that responded to auditory and visual speech was found in the posterior STS by using fMRI. TMS reduced the perception of the McGurk effect when a single pulse was delivered between 100ms before onset of the auditory stimuli to 100ms after onset of the auditory stimulus. *Conclusions:* Temporary disruption of the STS with TMS causes a significant reduction in the participants' perception of the McGurk effect. *Relevance to current work:* This study provided further evidence that the STS is a site of AV integration and is involved in the processing of the McGurk effect. The current study supports the observation that the STS is a region involved in auditory-visual (AV) integration. *Level of Evidence:* Level IIIa.

Bomba, M. D., Choly, D., & Pang, E. W. (2011). Phoneme discrimination and mismatch negativity in English and Japanese speakers. *Neuroreport, 22*(10), 479-483. doi: 10.1097/WNR.0b013e328347dada

*Objective:* The purpose of this study was to examine MMN differences in certain phonemes between English and Japanese speakers and to compare MMN of glides, liquids, and vowels in native and non-native English speakers. By examining these two components, the study overall examined how different types of vowels and consonant-vowel phonemes are processed in the brain. *Study Sample:* Sixteen adults participated in the study, eight native English speakers and eight native Japanese speakers who learned English after the age of 12 years. *Methods:* The stimuli consisted of vowels and consonant-vowel syllables presented in separate sequences. For all sequences, 1000 stimuli were presented randomly and consisted of a standard stimulus and a deviant stimulus. The standard stimuli were presented 85% of the time and the deviant stimuli were presented for the remaining 15% of the time. The Standard English vowel was /iy/ and the deviant vowel was /i/. The consonant-vowel syllables consisted of a standard set (/da/ and /ra/) and a deviant set (/wa/ and /la/). EEG was recorded for each participant from 26 electrodes. *Results:* When analyzing the vowel set there was no significant differences between native and non-native speakers on MMN latency and amplitude. In the /i/ versus /iy/ condition, there was a clear MMN along the frontal chain (F3, Fz, F4) and vertex (Cz) for both native and non-native speakers with a latency at about 200 ms. For the /da/ versus /wa/ syllables, MMN latency showed no significant differences between native and non-native speakers. For the native English speakers in the /da/ versus /wa/ condition, there was a MMN response along the frontal chain (F3, Fz, F4) and vertex (Cz) electrodes with a latency just greater than 200 ms. For the non-native speakers the MMN was smaller in amplitude along the frontal chain and almost impossible to identify at the Cz electrode. The /ra/ versus /la/ condition showed low amplitude MMN compared to the other stimulus conditions in native speakers. A clear MMN response was seen in 75% of the native participants with only 25% showing poor MMN-like responses. For the non-native speakers, the MMN was absent in 50% of the subjects. In 38% of the subjects the MMN had extremely low amplitudes. *Conclusions:* The results show that native English speakers had larger amplitude MMNs than the non-native speakers in the consonant-vowel syllable conditions. However, there was no difference in the MMNs in the vowel condition. Vowels and consonants were observed to be processed differently in the brain as measured by MMN. The differences found in the MMNs of the consonant-vowel syllable conditions between the two groups showed that neural differences exist in phonemic processing between speakers of different languages depending on the level of exposure of a particular phoneme. *Relevance to Current Work:* The previous study indicated that phonemes are perceived differently based on a person's native language. This evidence suggested that participants in the current study needed to be native English speakers in order to avoid differences that may arise in the data due to language knowledge and memory. *Level of evidence:* Level IIIa.

Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society*, *363*, 1001–1010. doi: 10.1098/rstb.2007.2155

*Objective*: This paper is a selective review about a variety of ways that the visual input from the speaker influences the auditory perception of speech. *Conclusions*: The review begins with the source-filter model and a discussion of the physical characteristics of the system. This segues into the topic of binding and the McGurk effect. Next, the contribution of vision (e.g. speech reading) was addressed. The research of Auer and Bernstein (1997) was cited where it was observed that 12 phonemically equivalent classes were enough to identify most English words. The most efficient speech reading occurs when mouth opening and closing, as well as tongue position, are clearly visible. Visible movements provide critical components of the AV advantage as the audible and visual patterns of speech are highly correlated; thus, AV processing is more effective than auditory only processing of natural speech. Calvert et al. (2004) and Bernstein et al. 2002 reported that the left hemisphere is activated during silent speech reading. Usually the right hemisphere is activated more when gazing at facial expressions. During visual speech detection tasks, the posterior STS (pSTS) has been found to be activated during natural and sequenced still-image conditions though the pSTS is more strongly activated when processing normal movement. Still-frame images of lips have not been shown to activate the pSTS (Capek et al., 2005). *Relevance to current work*: The research done in this study played an important role in the current study because it concluded that a combined AV processing is most effective in processing natural speech and that the pSTS is activated during this process. *Level of Evidence*: Level 1.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology, 113*, 495-506. doi:10.1016/S1388-2457(02)00024-X

*Objective:* This study was designed to assess the existence of an MMN evoked by the McGurk effect with constant auditory components and to support the hypothesis behind the revised motor theory that a rare, incongruent visual stimuli dubbed onto constant auditory syllables will evoke an MMN by creating a deviant phonetic percept through the McGurk effect. *Study Sample:* Eight native French speakers (ages ranging from 17-62 years) with clear MMN evoked by pure tones were selected to participate in this study. *Methods:* Five experiments were conducted. Experiment 1 targeted MMN evoked only by auditory stimuli; the video screen was off. Experiment 2 targeted MMN evoked only by visual stimuli; the sound was off. Experiment 3 targeted MMN evoked by AV stimuli, showing both congruent and incongruent stimuli. For each of these experiments, the same contrast was never presented more than once in immediate succession, and the same modality was never presented more than twice in a row. Experiment 4 focused on the inversion of polarity in AV MMN. Only one subject was used in this particular experiment. The four contrasts investigated were pure tone frequency contrast, auditory spatial localization contrast, auditory syllable contrast, and AV syllable contrast evoking the McGurk effect. Experiment 5 measured the percentage of McGurk illusions perceived by each subject based on their answers on a multiple choice sheet. During experiments 1-4, subjects performed a tactile discrimination task. *Results:* Experiment 1. Clear MMNs were evoked at $F_z$ but not at $O_z$. Both of the auditory contrasts evoked an MMN that inverted its polarity between $F_z$ and $M_1$ as well as $M_2$ electrodes. Experiment 2. No MMN could be detected either at $F_z$ or at $O_z$. Experiment 3. Both phonetic contrasts created by the McGurk effect evoked an MMN at $F_z$, but not at $O_z$. Neither of the audiovisual contrasts evoked a significant MMN that inverted its polarity between the $F_z$ and $M_1$ or $M_2$ electrodes. Experiment 4. No statistically significant positivity was found at either mastoid electrode locations. Experiment 5. AV incongruent stimuli caused combination-type illusions in 74% of the trials and fusion-type illusions in 66% of the cases. *Conclusions:* Not all evoked MMN invert their polarity in spite of the production of significant negative waveforms. The authors suggest that the McGurk effect results from an automatic, pre-cognitive comparison between short term memory phonetic traces. *Relevance to current work:* The research done in this study played an important role in the current study because it concluded that MMN found in the McGurk effect occurs precognitively. *Level of Evidence:* Level IIIa.

Csépe, V., Osman-Sági, J., Molnár, M., & Gósy, M. (2001). Impaired speech perception in aphasic patients: Event-related potential and neuropsychological assessment. *Neuropsychologia, 39*(11), 1194-1208. doi: 10.1016/s0028-3932(01)00052-5

*Objective:* The purpose of this study was to evaluate whether the MMN response to auditory stimuli (speech and non-speech) was deviant in individuals with aphasia. Also, the study aimed to determine whether impairment was due to a deficit in phonemic processing or related to phonetic features. Overall, the study evaluated how aphasic individuals processed language and how it deviated from typical processing. *Study Sample:* Four diagnosed aphasic patients and four neurologically unimpaired, control participants took part in this study. *Methods:* Three different types of stimuli were presented to the participants. The stimuli consisted of pure tones, front vowels, and consonant-vowel (CV) syllables. Event-related potentials were recorded for each individual with 21 electrodes using Neuroscan software. The ERPs were collected as the participants were presented auditory stimuli. *Results:* In all control subjects a reliable MMN was collected for all stimuli. The four aphasic participants all had MMN abnormalities. Specifically, the MMN elicited by pitch deviations was not significant enough to distinguish between the aphasic patients and the control group. The MMN elicited by consonant contrasts proved to show the most significant difference in aphasic patients in comparison with the control group. Lastly, a significant difference was seen in the MMN elicited by voicing and place of articulation. Aphasic participants showed great anomalies in this MMN compared to the control group. The MMN collected from the aphasic participants was either limitedly distributed, distorted, or completely missing. *Conclusions:* This study concluded that the MMN elicited by contrasting features reflects deficient processes due to damaged or disconnected regions of the language-processing network seen in those with aphasia. MMN responses collected in individuals with aphasia were clearly deviant compared to those with unimpaired neurological systems; thus, demonstrating the affect that neurological damage has in brain processing. *Relevance to current work:* This research demonstrated the difference in brain processing that occurred when an individual was affected by neurological impairment. This demonstration supported the current study's exclusion of individuals with known neurological, cognitive, or learning impairments as it would have an impact on the accuracy of data. *Level of evidence:* Level IIIb.

Gentilucci, M., & Cattaneo, L. (2005). Automatic audiovisual integration in speech perception. *Experimental Brain Research, 167*(1), 66-75. doi: 10.1007/s00221-005-0008-z

*Objective:* This study was two-fold: to determine whether features of both the visual and acoustic inputs are always merged into the perceived representation of speech and whether this AV integration is based on either cross-modal binding functions or on imitation. *Study Sample:* Sixty-five right-handed Italian speakers (22-27 years) participated in this study. All were naïve to the purpose of the study as well as to the McGurk paradigm. *Methods:* Three experiments were conducted, each with a different set of participants. Experiment 1. The participants first silently read the string of letters before repeating them aloud. Experiment 2. Congruent and incongruent AV stimuli were presented, following the McGurk paradigm (mouth mimics /aga/ sound production). Experiment 3. Congruent and incongruent AV stimuli were presented, following the inverse McGurk paradigm (mouth mimics /aba). All participants were required to repeat aloud what they had perceived. Their lip movements were recorded using a 3D-optoelectronic ELITE system. At the end of each session, they filled out a questionnaire indicating whether the sound of each string of phonemes (i.e. /aba/, /ada/, and /aga) varied and whether they noticed any incongruence between the acoustical and the visual stimulus. *Results:* Experiment 1. The voice spectrum and the lip kinematics varied according to the pronounced strings of phonemes. Experiment 2. None of the participants reported noticing any incongruence between AV stimuli. The voice spectra recorded in the incongruent AV presentation was compared with those recorded in the congruent AV presentation. F2 of /aba/ repetition in the incongruent AV presentation was influenced by the visually presented /aga/. No significant difference was found between the lip kinematics of the two /aba/ pronunciations and between the lip kinematics of the two /ada/ pronunciations. Experiment 3. F2 of /aga/ pronounced in the incongruent AV presentation significantly decreased while F2 of /aba/ pronounced in the incongruent AV presentation significantly increased as compared to F2 of the same strings of phonemes pronounced in the congruent AV presentation. *Conclusions:* The participants perceived a different sound of the same string of phonemes (F2 shifted in direction of F2 of the string of phonemes presented in the other sensory modality) rather than perceiving a completely different string of phonemes. Only the kinematics of labial consonant pronunciation of the presented string of phonemes was extracted from the visual stimulus and integrated with the acoustic stimulus. The data support cross-modal integration hypothesis between the AV inputs, rather than superimposition of automatic imitation motor programs of acoustically on visually detected motor patterns. *Relevance to current work:* The research done in this study played an important role in the current study because it concluded that participants may hear a different variation of the phoneme, but not a large enough of a difference to change the perception/category of the sound they hear. While there may be variation in the perceived sound of the three selected phonemes in the present study, the listeners would be able to identify the correct category. *Level of Evidence:* Level IIIa.

Green, K. P., Kuhl, K. P., Meltzo, N. A., & Stevens, E. R. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics, 50*, 524-536.

*Objective:* This study examined the effect of a discrepancy between the gender of the talker of the auditory and visual signals, thus manipulating the cognitive congruence between the signals. *Study Sample:* A total of 88 subjects participated in this study. They were either paid or given course credit for their participation. All were native English speakers, had normal or corrected-to-normal vision, and had no history of a speech or hearing disorder. *Methods:* Experiment 1 examined the effect of a cross-gender discrepancy on the perception of the McGurk effect. The entire head of the speaker, face and hair, was recorded. The speakers were recorded saying /ba/ and /ga/, which clip was low-pass filtered at 9.89 kHz. Two of the stimuli were matched by gender while two types of stimuli were created by cross-dubbing AV information onto the other speaker. Two of the four AV stimuli had conflicting phonetic information (i.e. auditory /ba/ paired with visual /ga/). Each block of trials contained 40 trials, ten repetitions of a set of four AV stimuli with only one type of the male/female stimuli. A between-subjects design was selected; each participant randomly was assigned to either the female or male face condition. A smaller sample was assigned to either visual only or auditory only condition. Experiment 2 was a follow-up experiment to test to see if the results were tied to the particular faces and voices used in the first experiment. Two new talkers were selected and experiment 1 was repeated with new listeners. Experiment 3 directly assessed the influence of any discrepancy between the auditory and visual signals that could be detected by the subjects. Stimuli from the previous two experiments were used. A new set of listeners used a 10-point scale to measure how well-matched the signals were, with 10 signifying a perfect match between AV stimuli. They responded verbally in order to maintain their attention on the monitor. *Results:* Experiment 1. For the fusion stimuli, no reliable differences or interactions were found between the two factors (male and female voice and face). An effect was found on the voice factor; the female's voice produced more /g/ responses than the male voice did regardless of the face that was shown, meaning fewer combination responses were selected. Experiment 2. These results closely replicate the findings of experiment 1 and support the conclusion that the McGurk effect is not influenced by the incongruence of the stimuli with respect to the gender of the talker. Experiment 3. The listeners were able to correctly categorize the gender of the speaker and identify when incongruence were displayed. *Conclusions:* Experiment 1 showed that the McGurk effect is not affected by the gender of speaker. The effect is equally strong when the face-voice stimuli are gender-compatible or gender-incompatible. Experiment 2 confirmed the previous results, supporting the conclusion that the McGurk effect is not dependent on a particular talker or vowel (/a/ vs. /i/). However, the response pattern ("d" vs. "th") was significantly different for these vowels. The participants in Experiment 3 perceived the incompatibility in the cross-gender stimuli. Experiment 3 revealed that the participants were able to detect the discrepancy between cross-gender stimuli. This suggests that the results from the first two experiments were not attributed to the participants' inability to detect any incompatibility between the stimuli. *Relevance to current work:* This study found that differences in the gender of the talker do not impact the integration of the phonetic information. While the listener is aware of the details of the speaker, this information is neutralized for task of phonetic categorization. *Level of Evidence:* Level IIIa.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *NeuroReport, 14*, 1129-1133.

*Objective:* The purpose of this study was to evaluate the relationship between brain activation and the degree of AV integration of speech information during a phoneme categorization task in order to assess modulation effects directly related to perceptual performance. *Study Sample:* Twelve, right-handed participants (22-49 years) composed this study. *Methods:* The AV stimuli (/aba/ and /ava/) were presented either in synchrony or $\pm$ 400ms out of phase. Stimuli were presented 10 times in three blocks that lasted 30 seconds. To establish baseline, the participants looked at a static face for three, 30s blocks. This created a total of 21 blocks. The subjects were asked to identify whether they heard /b/ or some other consonant. fMRI was used to measure brain activity. *Results:* The highest amount of electrophysiological activation during incongruent stimuli presentation was located in the right supramarginal gyrus and left inferior parietal lobule. Activation of the right precentral gyrus was observed as well. *Conclusions:* The number of /b/ sounds reported by the subjects positively correlated with activation of the area near the occipital-temporal junction. Early presentation of auditory stimuli enabled auditory information to influence visual processing regions in the brain although this modulating mechanism is unknown. No relationship was found between perceptual performance and activation in the STS or auditory cortex. *Relevance to current work:* The research done in this study played an important role in the current study because it concluded that early presentation of the auditory portion of the stimuli had a stronger influence on the visual perception of the stimuli. Therefore, the current study presented the audio and the visual stimuli in synchrony. *Level of Evidence:* Level IIIa.

Kasai, K., Nakagome, K., Iwanami, A., Fukuda, M., Itoh, K., Koshida, I., & Kato, N. (2002). No effect of gender on tonal and phonetic mismatch negativity in normal adults assessed by a high-resolution EEG recording. *Cognitive Brain Research, 13*(3), 305-312. doi: 10.1016/s0926-6410(01)00125-2

*Objective:* The study was done in order to clarify the role of gender differences in auditory MMN by comparing the amplitude, latency, and topography of tonal and phonetic MMN. *Study Sample:* The experiment included 18 male participants and 10 female participants, all of whom were native Japanese speakers. *Methods:* Auditory ERPs were the index used to measure the MMN. The participants were presented with auditory stimulus sequences consisting of standard and deviant stimuli that were delivered randomly. The exception to the random pattern was that each deviant stimulus was preceded by at least one standard stimulus. The subjects were instructed to watch a silent film and were encouraged to ignore the stimuli. After the film, the subjects were required to report on the content of the film to ensure their attention was on the film. In addition they reported on the characteristics of the stimulus sequence to ensure they behaviorally perceived the duration of tones and the phoneme boundaries. The experiment looked at two conditions. First, it looked at the MMN in response to a duration change of pure-tone stimuli. Second, it looked at the MMN in response to an across-category vowel change. The EEG recording was done via a 128-electrode cap. The MMNs were measured using the difference waveforms obtained by subtracting the ERPs of standard stimuli from those of deviant stimuli. *Results:* The mean global field power peak latencies of the male and female groups were 162 ms for the pure-tone MMN and 156 and 170 ms for the phonetic MMN. These results indicated that there is no significant effect of gender on either pure-tone or phonetic MMN amplitude. The MMN topography also indicated that there were no differences between genders but there were differences between conditions. The latency of the MMN also did not show a difference between genders but showed a difference between conditions. The latencies of the pure-tone MMN were significantly longer compared to the phonetic MMN, but this was found in both genders. After the experiment, all participants reported they could concentrate on the film and reported the content of the film correctly. All of the participants also reported correct information about the stimuli they heard while watching the film. *Conclusions:* The experiment concluded that there is no effect of gender on the amplitude, latency, or topography of tonal and phonetic MMN in normal adults using EEG. The conclusion reached allows researchers to know that combining males and females in experiments will not have obscure effects. The study also concluded that the pure-tone MMN was generated from Heschl's gyrus, and the phonetic MMN was generated from the planum temporal. *Relevance to current work:* The current study used both male and female participants. The study summarized above observed that there is no gender difference in MMN measurements. This was important information to the current work because it confirmed that using both males and females in the study would not affect the collection or accuracy of the data. *Level of evidence:* Level IIIa.

Kislyuk, D. S., Mööttöönen, R., & Sams, M. (2008). Visual processing affects the neural basis of auditory discrimination. *Journal of Cognitive Neuroscience, 20*, 2175-2184. doi: 10.1162/jocn.2008.20152

*Objective:* The purpose of this study was to see if there would be a perceptual similarity, as measured by neural representations in the auditory cortex, between the standards and deviants in the McGurk condition. *Study Sample:* Eleven native Finnish speakers (mean age 25 years) participated in this study. *Methods:* The two auditory stimuli, /ba/ and /va/, were equalized for intensity. Two visual stimuli were presented: video showing a mouth clearly articulating /va/ and the other overlaid with a blue ellipse that would stretch and shrink in rhythm with the lips' movement. The acoustic stimulus was delayed by 110 msec after the onset of the visual movement. Three conditions were presented in a dim room: the auditory condition where a subtitled silent movie was played during the recording, the McGurk condition where the clip of a mouth clearly articulating /va/ was shown, and the ellipse condition where a pulsating ellipse covered an immobile mouth. Data was collected by recording the EEG and the participants' responses to identifying whether they heard /ba/ or /va/. *Results:* Data from the two participants who correctly identified more than half of the acoustic components of the McGurk stimuli were not included in the analyses as this indicates a weak McGurk effect. The remaining nine participants were all susceptible to the McGurk effect. Their ability to correctly identify /ba/ in the McGurk stimuli was significantly below chance. The auditory and ellipse conditions elicited similar MMN while the McGurk condition did not. The interaction between the stimulus type and the electrode location was significant in the auditory and ellipse conditions only. *Conclusions:* For those susceptible to the McGurk effect, their brain's processing of the same /ba/ syllable is affected by viewing speech, qualitatively changing the auditory percept at the auditory cortex level. Processing the visual speech may have modified the activity in the auditory pathway. This profoundly influences the auditory cortex mechanisms underlying early sound discrimination, suggesting multisensory interactions occur during early auditory processing. *Relevance to current work:* The research done in this study played an important role in the current study because the results further the evidence that conflicting signals from different modalities (e.g. auditory and visual) merge to form a unified neural representation during early sensory processing. This phenomenon was observed in the current study as well. *Level of Evidence:* Level IIIa.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*, 253-257.

*Objective:* This study assessed the existence of an MMN evoked by McGurk–MacDonald percepts elicited by AV stimuli with constant auditory components. *Study Sample:* This study included 44 participants between the ages of 18-24 years. *Methods:* A female speaker was recorded saying a series of CV utterances that contained either a stop plosive or a nasal /m, n/ with the vowel /a/. New recordings were made after dubbing of each CV syllables into reciprocal combinations (e.g. ba-lips/ga-voice; ga-lips/ba-voice). Four video films were created from a random combination of the 56 possible AV stimuli. Each film contained 22 trials, each separated by a 10 second presentation of a blank video tape. The order of presentation was randomized. Each sequence contained three repetitions of each AV composite and each series was separated by a 10 second presentation of a blank video tape. *Results:* Virtual interchangeability was found between the different places of articulation of nonlabial sounds (e.g. /da, ta, ga, ka, na/). These results confirm the manner-place hypothesis as the auditory presentation of these sounds did not elicit an illusion of the auditory stimulus. When labial sounds were combined with nonlabial lip movements, the mean error rate was 73% (range from 30-100%). Inversely, when nonlabial sounds were combined with labial lip movements, the average error rate was 25% (range from 0-75%). *Conclusions:* The results confirm the predictive validity of the manner-place hypothesis with regard to the illusions elicited by labial-voice/nonlabial lips presentations; the results were not as strong for nonlabial sound/labial lips. Thus, information about the visual place of articulation sometimes leads to an illusion, confirming the active, constructivist nature of the speech perception process. *Relevance to current work:* The research done in this study played an important role in the current study because it concluded that the presentation of the pairing of nonlabial and labial AV stimuli produced a strong influence on the visual perception of the stimuli. Therefore, the current study used labial and nonlabial audio and visual pairings. The current study adds additional understanding of the McGurk effect. *Level of Evidence:* Level IIIa.

Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: Evidence from articulatory suppression, the McGurk Effect, and fMRI. *Journal of Cognitive Neuroscience*. *26*(3), 606–620. doi: 10.1162/jocn_a_00515

*Objective*: The purpose of this study was to examine the role of the motor system in the integration of auditory and visual stimuli. *Study Sample*: A total of 50 (13, 17, and 20 per experiment, respectively) right-handed, native-English speakers participated in the study. M*ethods*: Experiment 1a: Participants listened to ten trials of each of the following stimuli: /pa/, /ta/, and /ka/ which were repeated four times in a row. Participants recorded their responses on an answer sheet. A low-amplitude, continuous white noise was added to mask other sounds. The secondary tasks included either continuously articulating, without voicing, the sequence "/pa/…/ba/" or continuously performing a finger-tapping sequence, 1-2-3-4-5-5-4-3-2-1 where 1 is the thumb and 5 is the pinky for the duration of the stimulus presentation. Experiment 1b: The set up was similar to experiment 1a, except the stimuli was presented once instead of four times, the stimulus duration was increased to 2000 ms, and the white noise level was increased from 10% to 20%. The participants silently articulated only /pa/ successively. Responses were recorded on a keyboard. Experiment 2: The stimuli in this experiment were similar to the previous experiment, except that the duration was 1000 ms and the noise level was increased to 25%. Visual only stimuli were added as well as an articulatory rehearsal condition. Each block consisted of 10 sequential identical speech sounds with a 2.5 sec interval separating the blocks. Data was collected using an fMRI machine. *Results*: Experiment 1a: Direct modulation of the listener's motor system via concurrent speech articulation did not modulate the strength of the McGurk effect; the McGurk effect was equally robust during both secondary tasks. Experiment 1b: The results were similar to 1a. McGurk fusion rate did not change from baseline during articulatory suppression. Experiment 2: Results analyzing the regions of interest suggest that AV integration for speech involves the pSTS but not the speech motor system. *Conclusions*: The results from these two studies suggest that integration of the AV information is not dependent on the activation of the motor system, providing evidence against the motor speech system's role in AV integration. No differences were found between the participants' responses during the articulatory suppression task and the finger-tapping task. Because the participants' congruent articulation was shown to have no effect, this suggests that the motor system is not part of the processing of the McGurk effect. *Relevance to current work*: This article provides more support that the pSTS is involved in the processing and integration of auditory and visual stimuli while providing further evidence against the involvement of the motor speech system in this process. *Level of Evidence*: Level IIIa.

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

*Objective*: The purpose of this study was to demonstrate the importance and effect of vision on speech perception. *Study Sample*: The study contained 103 participants (ages 3-40 years). They were divided into three age categories: preschool children (3-4 years), primary school children (7-8 years), and adults (18-40 years). *Methods*: A female was recorded saying /ba//ba/, /ga//ga/, /pa//pa/, and /ka//ka/. A half second pause was added in between repetitions. Four dubbed videos were created: voiced /ba/ and visual /ga/, the inverse, voiced /pa/ and visual /ka/. Each recording was composed of three of these pairs. Four sequences of recordings were made. A ten second gap separated each of these sequences. Each participant was tested under two conditions: AV where they repeated orally what they heard, and auditory only where they repeated what they had heard when their backs were to the screen. Each participant heard all four recordings in both conditions though the stimuli were presented in a different order. *Results*: Accuracy for the auditory only condition was between 91-97% for the three age groups. For the AV condition, the error rates were as follows: 59% for preschool subjects, 52% for primary school children, and 92% for adults. The voiced /ba/ and visual /ga/ pair elicited more fusion (e.g. /da//da/) than the voiceless pairs. Visual input had a larger impact in adults than in children, though all results for AV condition are statistically significant. *Conclusions*: When visual /ga/ and voiced /ba/ was presented, a fused result (e.g. /da//da/) was perceived most frequently. While there were age-related changes in susceptibility to the McGurk effect, this illusion was observed in young children as well as adults. *Relevance to current work*: This study is the foundational article to the current research. McGurk and MacDonald conducted this study in order to generalize their observation on the illusion that is known as the McGurk effect. The current study furthers the research on the McGurk effect. *Level of Evidence*: Level IIIa.

Möttönen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research, 13*(3), 417-425. doi: 10.1016/S0926-6410(02)00053-8

*Objective*: This study investigated whether change detection mechanisms in the auditory cortex can distinguish between phonetically different unimodal visual speech stimuli or if acoustic speech integration helps detect visual changes in the auditory cortex. *Study Sample*: Seven Finnish-speaking volunteers participated in the study. One was left-handed. Two of the original ten were excluded because they were not susceptible to the McGurk effect. *Methods*: A female Finnish speaker was recorded saying /ipi/, /iti/, and /ivi/. Stimuli consisted of congruent and incongruent pairs of each of the three sounds. In the visual experiment, only the visual stimuli were shown. In the AV experiment, the participants were asked to count the number of times /ivi/ was presented. The experiment consisted of three to four sessions that lasted 15 to 20 minutes each. In the visual experiment, participants followed a similar method, except there were only two sessions instead of three or four. MEG was used to measure brain activity. *Results*: Both congruent and incongruent deviants elicited reliable mismatch fields (MMFs) in the left hemisphere at 130-295 ms for congruent stimuli and 140-160 and 200-300 ms for incongruent stimuli after the onset of the acoustic stimuli. In the right hemisphere, reliable MMFs for both congruent and incongruent deviants were elicited. MMF to congruent deviant evoked larger peak amplitude. In the visual only experiment, reliable bilateral MMFs were elicited. In addition, these MMFs were delayed in comparison to the deviants in the AV experiment. *Conclusions*: This study found that changes in visual speech stimuli were detected in auditory cortices bilaterally even when acoustic stimuli were absent. In the visual only experiment, visual changes were processed at a longer latency implying that the integration of the AV information speeds up the processing rate in the auditory cortex. *Relevance to current work*: This study provides more evidence that AV speech may be integrated during the early stages of speech processing. The current study further examined this observation. *Level of Evidence*: Level IIIa.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58*, 351-362.

*Objective:* The purpose of this study is to clarify the influences of timing on AV integration on the McGurk effect. *Study Sample:* Sixty-three (15, 30, and 18 per experiment) native speakers participated in this study. Four of the nineteen potential participants for the first study were not included as they did not experience the McGurk effect. *Methods:* In all three experiments, the stimuli consisted of visual /aga/ or /igi/ that was paired with audio /aba/. Participants responded by selecting the key that best represented the sound they heard (e.g. /b/, /d/, /g/, or other). Experiment 1. A female was used to create the stimuli. The presentation of the auditory stimuli varied in steps of 60 ms from 360 ms pre-synchrony to 360 ms post-synchrony, creating 13 AV pairings per vowel. Ten blocks composed of the 26 AV pairings were randomly presented. Experiment 2. Three female speakers, who vary in the amount of facial motion, produced utterances in three different speaking conditions (i.e. fast, normal, and clear). Synchrony was maintained at the point of acoustic release. Nine pairs were created for each speaker for a total of 27 AV stimuli in each block. Experiment 3. The differences in the perception of the different AV combinations as a function of delay were analyzed. A subset of the AV speaking style combinations and timing conditions that were presented in the first two experiments (i.e. fast and clear productions, and delay of the auditory stimuli with respect to the video stimuli) was used. The release bursts in the stops were synchronized. In addition, the auditory stimuli were delayed by 50, 100, 150, 200, and 250 ms relative to the timing of the onset of the release burst. *Results:* Experiment 1. A significant effect for delay was found, with larger asynchronies producing more /b/ responses. More /b/ responses signify a weaker McGurk effect. For the vowel /a/, /b/ was reported less frequently when the auditory stimuli lagged behind the video signal by 60 ms rather than when the auditory signal was in synchrony with the sound track of the video signal. Experiment 2. The number of /b/ responses increased as the auditory speed shifted from fast to normal to clear. Inversely, the number of /b/ responses decreased almost in half as the visual speed shifted from fast to normal to clear. The number of /b/ responses also increased as the AV information became more dissimilar. Experiment 3. With the exception of the visual fast, auditory clear condition, the other three functions elicited a similar response pattern. This shows that the relative timing of the onsets or offsets of the bisyllables is not a big influencer on the McGurk effect and does not explain the results pattern observed in the previous experiment. *Conclusions:* A strict synchrony of the auditory and visual stimuli is not necessary to elicit the McGurk effect. However, the rates of visual and auditory stimuli have a significant influence on perception. A small, reliable tendency for the better matched stimuli to elicit more McGurk illusions than unmatched conditions was found. *Relevance to current work:* The research done in this study played an important role in the current study because it concluded the McGurk effect is most often observed when the AV stimuli are presented in synchrony. Therefore, the current study presented the audio and the visual stimuli in synchrony. *Level of Evidence:* Level IIIa.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing, 16*(1), 6-18. doi: 10.1097/00003446-199502000-00002

*Objective:* This article explained how the ERP component, specifically the MMN, can be used in understanding auditory function and forms of its pathology. Also, the article discussed how MMN can be used as an accurate objective measure in research. *Conclusions:* MMN is elicited when an acoustically deviant stimulus replaces a standard stimulus. The deviant stimulus creates a difference wave that is negative. This negativity that is created is generated by a change-discrimination process that occurs in the auditory cortex. The reason why MMN is a good measurement in research is that it is easy to elicit, provides an objective measure of discrimination ability, is elicited without attention, and central auditory representations are involved in MMN generation. In addition, MMN provides a representation of speech processing and reflects auditory sensory memory. The usefulness and properties of MMN is supported by numerous amounts of studies that the author has cited in the article. The author also proposed that MMN is elicited by two intracranial generators, one in the auditory cortex and the other in frontal areas. This proposal is also in harmony with other studies that have made the same discovery. *Relevance to current work:* The information obtained in this article provided a thorough definition of MMN. MMN played a major role in the collection of data in the current study. This article supported the use and analysis of MMN in the current research. *Level of evidence:* N/A.

Nahorna, O., Berthommier, F., & Schwartz, J. (2012). Binding and unbinding the auditory and visual streams in the McGurk effect. *Journal of the Acoustical Society of America, 132*, 1061-1077.

*Objective:* This study was designed to see if incoherent AV contexts can lead to unbinding of auditory and visual information, thus reducing the McGurk effect. *Study Sample:* Nineteen French subjects (between 22-27 years old) participated in experiment 1. Twenty French subjects (between 20-28 years old) participated in experiment 2. *Methods:* Experiment 1 was designed to decrease the McGurk effect by an incoherent AV context. Stimuli consisted of two parts: the "context", either coherent or incoherent, followed by the "target", either a congruent AV /ba/ syllable or an incongruent McGurk stimulus. The coherent context was made up of a sequence of 5, 10, 15, or 20 syllables. In the incoherent context, the only change was that the visual content was replaced by a series of random sentences matched in duration. The subject needed to identify whenever a /ba/ or /da/ syllable was heard. Experiment 2 was designed to test the role of phonetic vs. temporal incoherence in the McGurk modulation process. The stimuli were manipulated one of two ways: switching the auditory content from one syllable to the other and slightly advancing or delaying each auditory syllable between 30 ms to 170 ms. A similar procedure as experiment 1 was followed. *Results:* Experiment 1: An incoherent AV context at least five syllables long with a duration less than 4 seconds was sufficient to significantly decrease the McGurk effect. Experiment 2: Phonetic incoherence, and to a smaller degree temporal incoherence, increased the number of "ba" responses among most subjects. The duration of the context did not significantly affect the duration of incoherence. *Conclusions:* These experiments support the idea that McGurk fusion is dependent on the previous AV context and that unbinding can occur quickly. *Relevance to current work:* This study was important because it provided more information and ideas on how the McGurk effect continues to be studied and analyzed. *Level of Evidence:* Level IIIa.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage, 59*(1), 781-787. doi: http://dx.doi.org/10.1016/j.neuroimage.2011.07.024

*Objective:* The purpose of this study was to test the hypothesis that those who perceive the McGurk effect would have higher activity levels in their left STS when compared to those who do not perceive the McGurk effect, reflecting a lack of AV integration. *Study Sample:* Fourteen right-handed subjects (mean age 26.1 years) participated in this study. *Methods:* Stimuli were created by a female speaker and consisted of congruent (auditory and visual matching) syllables and two types of incongruent syllables (auditory and visual mismatch). Functional localizer scan series was used to identify the STS in each subject. Data were collected using both MRI and fMRI. For nine of the subjects the experiment consisted of 25 McGurk trials, 25 non-McGurk trials, 25 congruent /ga/ trials, 25 congruent /ba/ trials, 10 target trials (AV /ma/) and 30 trials of fixation baseline. *Results:* The McGurk susceptibility group had greater left STS activity to incongruent syllables than the non-perceivers. Across all subjects, there was a significant positive correlation between their STS response to incongruent syllables and their likelihood of experiencing the McGurk effect. No difference was found between these two groups for congruent syllables. *Conclusions:* The difference between McGurk perceivers and non-perceivers was found in the neural response in their left STS. The use of functional localizers to identify the location of the multisensory portion of STS in each individual sets this study apart from earlier similar studies. This study supports the idea that the STS is a critical brain locus for AV integration in speech perception. *Relevance to current work:* Therefore, the current study further examined the activation of the left STS. *Level of Evidence:* Level IIIa.

Neville, H. J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A., . . . Turner, R. (1998). Cerebral organization for language in deaf and hearing subjects: Biological constraints and effects of experience. *Proceedings of the National Academy of Sciences of the United States of America, 95*(3), 922-929. doi: 10.1073/pnas.95.3.922

*Objective:* The purpose of this study was to examine cerebral organization using fMRI in three groups of individuals with different language experiences. *Study Sample*: The three groups consisted of the following individuals: (a) normally hearing, monolingual, native English speakers who do not know American Sign Language (ASL), (b) congenitally, genetically deaf individuals who use ASL as their first language and later on learned English with no auditory input, and (c) normally hearing, bilingual subject with both ASL and English as native languages. *Methods:* All subjects from the aforementioned groups were right-handed, healthy adults. Each group of participants was scanned using fMRI while processing sentences in

English and ASL. The English sentences were presented on a screen and the ASL sentences consisted of a film of a signer producing the sentences. The materials were presented in four different runs, two English runs and two ASL runs. At the end of each run, participants were asked yes/no recognition questions on the stimuli to ensure attention. Images were collected for both the left and right hemispheres and comparisons of the images were made across hemispheres and languages. Also, regions of activation were observed and evaluated. *Results:* When normally hearing subjects read English sentences, activation was observed in the left hemisphere in areas including Broca's area, Wernicke's area, and the angular gyrus. Weak and variable activation was seen in the right hemisphere. On the other hand, deaf subjects did not display left hemisphere dominance when reading English and instead displayed activation in the middle and posterior temporal-parietal structures in the right hemisphere. When the monolingual hearing individuals viewed the ASL, they displayed no significant activation. However, when deaf subjects processed ASL there was a significant activation area in the left hemisphere within Broca's and Wernicke's area. Also, significant activation was identified in the right hemisphere. These results were also seen in the hearing individuals who also knew ASL as they viewed the ASL film. *Conclusions:* The study concluded that the processing of a person's native language occurs predominantly in the left hemisphere. This suggests that there are strong biological constraints that render particular areas in the left hemisphere of the brain to be designed to process linguistic information. *Relevance to current work:* This study found that native language is predominantly processed in the left hemisphere. The stimuli that were used in the current study consisted of phonemes from the English language. Because English was being used, it was vital that all participants were native English speakers because the above study suggests that the brain processes secondary languages differently. *Level of evidence:* Level IIIa.

Paré, M., Richler, R., Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception and Psychophysics, 65*, 553-567. doi: 10.3758/BF03194582

*Objective:* The purpose of this study was to examine the influence of gaze behavior and fixation on AV speech processing. *Study Sample*: Sixty-one participants (ages 18-35 years) took part in this study. *Methods*: Stimuli consisted of five nonsense utterances, /aba/, /ada/, /aga/, "atha", and /ava/, spoken by three females and one male. Each speaker was filmed in front of a blue background with only the speaker's head and shoulders in the frame. The subjects were seated at a desk with their heads restrained in a head- and chinrest which kept the subject's eyes 114 cm from a 20-in. television monitor. Responses were recorded on a keyboard labeled with the possible responses (b, th, v, d, g, and o for other). Experiment 1. The participants' gaze was monitored by a search-coil-in-magnetic-field technique during the presentation of 180 trials. The three regions consisted of the speaker's mouth and each of the eyes. Experiment 2. The participants were instructed to fixate on the speaker's mouth, eyes, or hairline, yielding a total of 324 trails (108 per gaze position). Experiment 3. The participants were instructed fixate on spots of lights that were beyond both the talker's head and the video monitor, for a total of 540 trials (135 for each of the 4 gaze positions). *Results*: Experiment 1. Each subjects' gaze behavior was typical for both the congruent and incongruent conditions; they fixated on only a few facial features. A statistically significant narrowing of gaze was observed in all subjects between the onsets of the first and second vowels. The majority fixated on the mouth (62%). Experiment 2. The main effects of gaze fixation position, stimulus and talker were statistically significant. No significant differences in the perception of the McGurk effect were found between gaze fixations on the mouth and the eyes. Experiment 3. The main effect of gaze fixation position was highly significant ($p < .001$). *Conclusions*: The perception of the McGurk effect was not significantly enhanced by fixating on the mouth which suggests that fixation in this region is not necessary for the integration of AV information to take place. The second experiment confirmed that fixating within the central region (eyes and mouth) of the speaker's face provides similar results when processing visual speech information. The third experiment showed that a significant McGurk effect could be produced when the subject's fixation deviated up to 40º from the talker's mouth. *Relevance to current work*: This study shows that the perception of the McGurk effect is effected if the viewer's gaze is fixated more than 40º from the speaker's mouth. The present study limited the areas potential gaze fixation by displaying only the mouth area. *Level of Evidence*: Level IIIa.

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. *Journal of Speech, Language, and Hearing Research, 52*, 1073-1081. doi: 1092-4388/09/5204-1073

*Objective:* The purpose of this study was test if the amplitude reduction effect in N1/P2 is actually associated with AV integration mechanisms or if it is attributed to another process. *Study Sample:* This study was composed of 24 participants (ages 18-30 years). Twelve people participated in Experiment A, and a different set of twelve people participated in Experiment B. *Methods:* A male speaker was recorded saying /pa/ and /ta/. Four examples of each syllable were normalized and calibrated at approximately 60 dB SPL. The first stilled frame was presented for 1,000 ms before the frames ran at 25 frames per second. In the AV condition, the AV stimui were presented in synchrony. In the AV asynchrony condition, the auditory stimulus was presented 200 ms in advance of the visual presentation. In the auditory only condition, the participants were shown a static fixation cross. In the visual only condition, the auditory stimuli were not played. EEG was used to record the results. *Results:* Between the auditory only and AV conditions, the AV conditions had significantly lower response amplitudes in both experiments. The maximal difference was recorded at Cz. Amplitude reduction was not found during the AV asynchrony condition. *Conclusions:* The peak wave of the N1/P2 wave following the presentation of AV speech was significantly smaller in comparison to the auditory only stimuli as well as the sum of the unimodal responses. This shows that the effect of AV speech was nonlinear. In order for an amplitude reduction to occur, the stimuli needed to be in synchrony. This supports the notion that this amplitude reduction effect is linked with the operation of integrative mechanisms and that some integration of auditory and visual information take place at an early stage. In addition, this study adds evidence to support the belief that the STS is a site of AV speech integration and is the source of inhibiting effects in the auditory cortex. *Relevance to current work:* The research done in this study played an important role in the current study because it adds additional evidence that the STS is involved in AV integration. Therefore, the current study further examined the activation of STS. *Level of Evidence:* Level IIIa.

Ponton, C. W., Bernstein, L. E., and Auer, E. T. (2009). Mismatch negativity with visual-only and audiovisual speech. *Brain Topography*, *21*, 207-215. doi:10.1007/s10548-009-0094-5

*Objective*: The purpose of this study was to further investigate the visual speech MMN by examining visual only and AV conditions. Current density reconstruction (CDR) models were also computed. *Study Sample*: Twelve, right-handed adults (20-37 years) participated in this study. *Methods*: The purpose of the study was explained to the participants. An initial screening confirmed that all participants were susceptible to the McGurk effect. EEG was used to measure brain activity during presentation of stimuli in a MMN paradigm where 87% were standard trials and 13% were deviant trials. Each participant was presented with a total of 4400 trials. Testing lasted approximately 4.5 hours. Horizontal and vertical eye movements were recorded on two differential recording channels. Recording began 100 ms before the acoustic onset and ended 500 ms post onset. The stimulus was held constant when MMN was calculated. While CDR models were examined, the interpretation of the MMN activity did not focus primarily on the CDR dipole. *Results*: The MMN for visual only /ba/ showed an increase in activity centered at 82 ms. The MMN for visual only /ga/ recorded the highest amplitude activity at 161 ms, a much longer latency. Unlike visual only /ba/, visual only /ga/ did not result in a unique prominent region of increased activity. The distributions of activity were observed to be less stable with MMN time waveforms than with the integrated MMN waveforms. Neither MMN nor integrated MMN waveforms resulted in left-hemisphere activity. CDR models based on time waveforms resulted in temporal lobe activity that was centered in the inferior temporal gyrus. *Conclusions*: CDR modeling was reliable for visual only and congruent AV /ba/. The CDR models showed that right lateral middle to posterior temporal cortex region was activated at short duration latencies in response to the presentation of visual only and congruent audio-visual stimuli. This finding suggests this temporal region plays a role in the representation of visible speech. In addition, the latencies of the MMNs obtained in this study were earlier than the latencies reported in other studies with regard to integrative AV effects in classical temporal auditory areas (e.g., Möttönen et al. 2002; Saint-Amour et al. 2007; Sams et al. 1991). *Relevance to current work*: This study used EEG to measure MMN in their research successfully. The current study also used EEG as an objective measure of brain activity. *Level of Evidence*: Level IIIa.

Saint-Amour, D., Sanctis, P. D., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*, 587-597. doi: 10.1016/j.neuropsychologia.2006.03.036

*Objective:* The purpose of this study was to further characterize the McGurk MMN by using high-density electrical mapping and source analysis to further examine the underlying cortical sources of this activity. *Study Sample:* This study was composed of eleven subjects (19-33 years). Post-study debriefing was conducted in order to ensure that all experienced strong McGurk illusions. *Methods:* A male speaker was recorded saying /ba/ and /va/. A spoken /ba/ was dubbed onto the video recording of /va/ to create an illusory McGurk AV pairing. Two conditions were presented (congruent AV /ba/; incongruent auditory /ba/ and visual /va/). Blocks lasted approximately one and a half minutes in order to minimize fatigue. Visual and AV blocks (35–40 stimuli/block) were randomly presented and separated by short breaks for a total of approximately 1420 trials per condition. The vision only condition was used in order to rule out a

McGurk MMN attributed to visual mismatch processes. This was done by subtracting the standard and deviant visual responses from the corresponding AV responses. Data were collected by running a continuous EEG. Topographical mapping and source localization, using Brain Electrical Source Analysis software, were also conducted. *Results:* There was no visual MMN for the visual alone condition. A robust MMN response was found in the latency range from 175-400ms. A significant main effect for hemisphere, showing left lateralization, was found for the initial phase of the MMN. The use of three dipoles accounted for 85% of the variance in the data. Right hemispheric contributions were accounted for with a single source in the STG and two separate sources were found in the left hemisphere (in the transverse gyrus of Heschl and in the STG). *Conclusions:* Visually driven multisensory illusory phonetic percepts are associated with an auditory MMN cortical response. The left hemisphere temporal cortex is important in this process. In addition, this study observed that the visual stimuli influenced auditory speech perception in the auditory cortex. *Relevance to current work:* This study found further evidence to support the notion that left hemisphere temporal cortex plays a crucial role in phonetic processing. Therefore, the current study further examined the activation of the left hemisphere. *Level of Evidence:* Level IIIa.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S .T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*, 141–145. doi: 10.1016/0304-3940(91)90914-F

*Objective:* The purpose of this study was to identify the neuroanatomical area where AV integration occurs, using MEG recordings. *Study Sample:* Ten adults participated in the study. *Methods:* The stimuli consisted of a Finnish female saying /pa/ and /ka/. The stimuli were concordant 84% of the time while the stimuli were presented in discordance the remaining 16% of the time. In some of the subjects, this probability was reversed. The cerebral source was modelled with an equivalent current dipole (ECD). *Results:* Starting around 180 ms, the waveforms elicited by the discordant and concordant stimuli began to differ. Visual stimuli, shown without the auditory component, elicited no response over the left temporal area in the two subjects that were studied. ECDs explained 95-96% of the field variance for the frequently presented stimuli in each set. The values for the minority stimuli in each set were 91% and 88% for the discordant and concordant respectively. The ECDs for both the difference waveforms and responses at 100 ms were oriented downwards. *Conclusions:* Visual information from articulatory movements can enter the auditory cortex and influence perception. *Relevance to current work:* This study provides more evidence that visual information does indeed affect auditory perception and is processed in the auditory cortex. The current study also shows that cognitive differences can begin in the early latency epoch. *Level of Evidence:* Level IIIa.

Szycik, G., Stadler, J., Tempelmann, C., & Munte, T. (2012). Examining the McGurk illusion using high-field 7 tesla functional MRI. *Frontiers in Human Neuroscience, 6*, 1-7. doi: 10.3389/fnhum.2012.00095

*Objective:* This study was designed to analyze brain sites that are involved in the processing and fusion of speech, especially when incongruent visual information is presented. *Study Sample:* Twelve right-handed, native speakers (21-39 years) participated in the study. *Methods:* The AV stimuli presented were /gaga/, /baba/, /dada/, and one syllable pair designed to cause the McGurk illusion (mismatched /dada/). The participants were instructed to press one button when they perceived /dada/ and another button when they heard any other syllable pair. 120 stimuli were presented. A Tesla7 fMRI machine was used, focusing on the frontal speech areas and the middle and posterior part of the STS. AV mismatched /dada/ events that caused the McGurk illusion were compared to AV mismatched /dada/ events that did not lead to the illusion. *Results:* The left and right insula showed greater activity for the McGurk illusion of /dada/ in comparison to the spoken stimulus /dada/. Differences in brain responses to the incongruent /dada/ stimuli were found between the illusion and the non-illusion group. Those in the illusion group had greater activation of the STS. *Conclusions:* This study suggests that the bilateral STS region is a major site for AV integration and that the left STS is a key area for individual differences in speech perception. *Relevance to current work:* This particular study found that auditory and visual integration occurs in the STS. The current study found further evidence to generally support this finding. *Level of Evidence:* Level IIIa.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*, 989-994. doi: http://dx.doi.org/10.1016/S0028-3932(02)00316-0

*Objective:* The purpose of this study was to examine whether visual perception of speech might also modulate motor excitability in the speech production system using transcranial magnetic stimulation (TMS). *Study Sample:* Fifteen subjects (19-40 years) participated in this study. *Methods:* After training the subjects for ten minutes on how to produce a constant level of contraction on the lip muscles, surface electrodes were attached to their orbicularis oris muscle to record EMG activity. Similar procedures were followed for the hand experiment, with the exception that the electrodes were attached to the first dorsal interosseous muscle in the right hand. The coil was arranged so that the current flowed in a posterior to anterior direction. The four experimental conditions were a speech condition (listening to speech while viewing visual noise), a non-verbal condition (listening to non-verbal sounds, e.g. glass breaking, bells ringing, guns firing, while viewing visual noise), a lips condition (viewing speech-related lip movements while listening to white noise), and an eyes condition (viewing eye and brow movements while listening to white noise). *Results:* No significant difference was found between the two hemispheres when the active motor thresholds for stimulation over the face area of the primary motor cortex were averaged. The main effects of hemisphere and condition were found to be significant, but the interaction between hemisphere and condition was not. Following left hemisphere stimulation, the motor-evoked potential ratios were greater than 100% for three of the four conditions. However, these ratios were less than 100% for all four conditions after right hemisphere stimulation. This indicates that none of the conditions increased motor excitability relative to the control condition following right hemisphere stimulation. No significant difference in the baseline EMG activity was found between the hemispheres. In the hand experiment, no significant difference was found. *Conclusions:* Speech perception, by listening to speech as well as by viewing speech-related movements, showed increased excitability of the motor units associated with speech production especially in the left hemisphere. This suggests that the left hemisphere may be specialized for imitation. *Relevance to current work:* This study found further evidence that the left hemisphere is associated with speech processing. Therefore, the current study paid particular attention to left-hemisphere activation. *Level of Evidence:* Level IIIa.

World Medical Association (2008). WMA declaration of Helsinki: Ethical principles for medical research involving human subjects. World Medical Association, Inc. Retrieved from http://www.wma.net/en/30publications/10policies/b3/index.html

*Objective:* This document was created by the World Medical Association (WMA) as a statement of ethical principles that should be followed for medical research involving human subjects. The principles also should be followed in research involving identifiable human material and data. *Relevance to current work:* The current study was done in an ethical manner in harmony with the principles stated in the Declaration of Helsinki. Along with being in accordance with the Declaration of Helsinki, the current research was also conducted under the ethical principles upheld by Brigham Young University's Institutional Review Board (IRB). In addition, the current study was approved by the Brigham Young University IRB. *Level of evidence:* N/A.

**Appendix B**

**Informed Consent to Act as a Human Research Subject**

Brain Mapping of the Mismatch Negativity Response of the McGurk Effect in Musical Performance and Visual Arts Students

David L. McPherson, Ph.D.
Communication Science and Disorders
Brigham Young University
(801) 422-6458

Name of Participant: ______________________________________

**Purpose of Study**
The purpose of the proposed research project is to study whether specific locations of brain activity are influenced more powerfully by visual or auditory stimuli. Also, the research project attempts to investigate the temporal resolution differences between auditory and visual speech processing.

**Procedures**
You have been asked to participate in this study by Lauren Nordstrom, B.S., a student conducting research under the direction of Dr. David L. McPherson. The study will be conducted in room 110 of the John Taylor Building on the campus of Brigham Young University as well as in 155 McDonald Building. The testing will consist of one to two sessions including orientation and testing and will last for no more than 3 hours. You may ask for a break at any time during testing. Basic hearing tests will be administered during the first half-hour of the session.

Surface electrodes (metal discs about the size of a dime) will be used to record electrical activity of your brain. These discs will be applied to the surface of the skin with a liquid and are easily removed with water. Blunt needles will be used as a part of this study to help apply the electrode liquid. They will *never* be used to puncture the skin.

Acoustic and linguistic processing will be measured using an electrode cap, which simply measure the electrical activity of your brain and *does not* emit electricity; no electrical impulses will be applied to the brain. These measurements of the electrical activity are similar to what is known as an "EEG" or brain wave testing. These measurements are of normal, continuous electrical activity naturally found in the brain.

You will wear the electrode cap while you listen to different syllables, during which time the electrical activity of your brain will be recorded on a computer. The sounds will be presented through speakers at a comfortable, but not loud listening level. You will be seated comfortably in a sound treated testing room. You will be asked to give responses during the hearing test and portions of the electrophysiological recording by pressing a series of buttons.

The procedures used to record the electrophysiological responses of the brain are standardized and have been used without incident in many previous investigations. The combination of sounds presented is experimental, but the recording procedure is not.

A structural MRI of the head will be obtained using standard clinical protocol and in accordance with guidelines and policies of the BYU MRI Research Facility. You will complete the BYU MRIRF Screening Form immediately preceding the scan in order to ensure you meet the safety requirements to use the equipment.

**Risks/Discomforts**
There are very few potential risks from these procedures, and these risks are minimal. The risks of using an EEG include possible allergic reactions to the liquid used in applying the electrodes. Allergic reactions to the liquid are extremely rare. There is also a possibility for an allergic reaction to the electrodes. If any of these reactions occur, a rash would appear.

Treatment would include removing the electrodes and liquid and exposing the site to air, resulting in removal of the irritation. If there is an allergic reaction, testing procedures would be discontinued. Another unlikely risk is a small abrasion on the scalp when the blunt needle is used to place electrode gel. Treatment would also include removing the electrode and gel, exposing the site to air and testing procedures would be discontinued.

There are very few potential risks from using an MRI as images are formed without x-ray exposure. Participants should be free of any metallic materials (i.e. artificial joints, metallic bone plates, heart pace maker, insulin pumps, etc.) as the magnets may joggle the metal inside of their bodies. Some participants may experience a claustrophobic sensation during the MRI. Before entering the MRI, the participants will be instructed to relax and to breathe normally while inside the machine. The MRI staff will be nearby during the scan. Treatment would include terminating the scan if the sensation becomes too great.

**Benefits**
You will receive a copy of your hearing assessment at no charge. You will be notified if any indications of hearing loss are found in this area. The information obtained from this study may help to further the understanding of language processing, which will be beneficial to professionals involved in treating speech and hearing disorders.

**Confidentiality**
All information obtained from testing is confidential and is protected under the laws governing privacy. All identifying references will be removed and replaced by control numbers. Data collected in this study will be stored in a secured area accessible only to personnel associated with the study. Data will be reported in aggregate form without individual identifying information.

**Compensation**
You will be given $10.00 compensation at the end of the session; you will receive this compensation whether or not you complete the study.

**Participation**
Participation in this research study is voluntary. You have the right to withdraw at any time or refuse to participate entirely without affecting your standing with the University.

**Questions about the Research**
If there are any further questions or concerns regarding this study, you may ask the investigator or contact David McPherson, Ph.D, Communication Science and Disorders, at (801) 422-6458; Taylor Building Room 129, Brigham Young University, Provo, Utah 84602; e-mail: david_mcpherson@byu.edu.

**Questions about your Rights as a Research Participant**
If you have questions regarding your rights as a research participant, you may contact the BYU IRB Administrator at (801) 422-1461; A-285 ASB, Brigham Young University, Provo, UT 84602; e-mail: irb@byu.edu.

**Other Considerations**
There are no charges incurred by you for participation in this study. There is no treatment or intervention involved in this study.

The procedures listed above have been explained to me by: ______________________________ in a satisfactory manner and any questions relation to such risks have been answered.

I understand what is involved in participating in this research study. My questions have been answered and I have been offered a copy of this form for my records. I understand that I may withdraw from participating at any time. I agree to participate in this study.

Printed Name:___________________________

Signature:______________________________

Date:__________________________________